

Ein prozessorientiertes Auswertungsmodell zur szenario-unspezifischen Einschätzung und Klassifizierung von Problemlöseleistung

Abhandlung
zur Erlangung der Doktorwürde
der Philosophischen Fakultät
der
Universität Zürich

vorgelegt von
Simone Denise Taner-Meister
von Zürich

Angenommen im Herbstsemester 2007 auf Antrag von
Herrn Prof. Dr. René Hirsig und Herrn Prof. Dr. Vinzenz Morger

Zentralstelle der Studentenschaft der Universität Zürich
Zürich, 2007

Inhaltsverzeichnis

Zusammenfassung	IV
Danksagung	VI
1 EINLEITUNG	1
2 THEORETISCHER HINTERGRUND	7
2.1 Ansätze zur Erfassung komplexer Problemlöseleistungen	7
2.1.1 Erfassung der Steuerungsleistung	8
2.1.2 Erfassung kognitiver Aspekte der Problembearbeitung	10
2.1.3 Erfassung von Verhalten und Strategien	16
2.2 Problemlöseszenarien als eignungsdiagnostische Instrumente	20
2.2.1 Begriffsdefinition: Personalauswahl und Managementdiagnostik	21
2.2.2 Vorteile und Nutzen der Szenarien	25
2.2.3 Qualitätsanforderungen und ihre Überprüfung	28
2.2.3.1 Überprüfung der Reliabilität	29
2.2.3.2 Überprüfung der Validität und die Intelligenzdebatte	31
2.2.3.3 Mögliche Ursachen für die heterogene Befundlage	35
2.3 Folgerungen und die Beschreibung eines neuen Auswertungsmodells	41
2.4 Fragestellungen und Hypothesen	47
3 METHODISCHES VORGEHEN	50
3.1 Anwendungsbeispiel: Das Szenario ATLANTIS	50
3.1.1 Der simulierte Realitätsausschnitt	51
3.1.2 Darstellung des Spielverlaufs	53
3.1.3 Beschreibung des bestehenden Auswertungsansatzes	58
3.1.3.1 Suchweg und Suchstrategie	58
3.1.3.2 Protokollbogen mit Leistungskennwerten	60
3.1.3.3 Anwendung der Auswertung in der Praxis	62
3.2 Entwicklung des neuen Auswertungsmodells	66
3.2.1 Prozessorientierte Mikroverhaltensbeschreibung	66
3.2.1.1 Datenbasis und allgemeine Definitionen	66
3.2.1.2 Kriterien für das Informationseröffnende Handeln	71
3.2.1.3 Kriterien für das Adäquate Handeln	73
3.2.1.4 Inhalt der verfassten Texteinheiten	77
3.2.2 Prozessdarstellung und -beschreibung	81

3.2.3	Beurteilung der Ähnlichkeit der Zeitreihen	84
3.3	Beschreibung des Expertenurteils	86
3.4	Vorgehen und Stichprobe	87
3.5	Auswertungsverfahren	89
4	ERGEBNISSE UND ERSTE INTERPRETATION	93
4.1	Repräsentativität der Teilstichproben.....	93
4.2	Darstellung und Beschreibung der Zeitreihen	95
4.2.1	Qualitative Beschreibung	96
4.2.2	Quantitative Beschreibung.....	101
4.3	Definition von Prozessprofilen mittels Klassifizierung der Zeitreihen	104
4.3.1	Klassifizierung aufgrund inhaltlicher Überlegungen.....	104
4.3.1.1	Prozessprofile für das Informationseröffnende Handeln	105
4.3.1.2	Prozessprofile für das Adäquate Handeln	106
4.3.2	Klassifizierung aufgrund formaler Methoden.....	109
4.3.2.1	Anmerkungen zu den verwendeten Methoden	109
4.3.2.2	Lösung und Interpretation für das Informationseröffnende Handeln	112
4.3.2.3	Lösung und Interpretation für das Adäquate Handeln	117
4.3.2.4	Zusammenfassung und Diskussion	121
4.3.3	Beurteilungsschema und Problemlösetypen: Beschreibung und Diskussion	123
4.4	Überprüfung des Auswertungsmodells.....	127
4.4.1	Vergleich mit der szenariospezifischen Auswertung	128
4.4.2	Vergleich mit dem Expertenurteil.....	133
4.4.3	Überprüfung an einem unabhängigen Datensatz	135
5	DISKUSSION UND AUSBLICK	138
6	LITERATURVERZEICHNIS	154
7	ANHANG	161
Anhang A:	Abbildungs- und Tabellenverzeichnis	162
Anhang B:	Ergänzungen zum Szenario ATLANTIS.....	164
Anhang C:	Ergänzungen zur Mikroverhaltensbeschreibung	174
Anhang D:	Ergänzungen zu den Ergebnissen	181

Zusammenfassung

Computergestützte Problemlöseszenarien bieten die Möglichkeit, individuelle Kompetenzen im Umgang mit komplexen, dynamischen Problemen zu diagnostizieren. Diese Arbeit präsentiert ein neues Auswertungsmodell, das die praktische Relevanz der Szenarien im Rahmen der Personalselektion erhöhen soll, indem Datenauswertung, -interpretation und -integration standardisiert werden. Resultat der neuen Auswertung ist ein vom Inhalt des jeweiligen Szenarios unabhängiges, mikroverhaltensbasiertes, prozessorientiertes Gesamturteil über die gezeigte Problemlöseleistung. Das Gesamturteil besteht aus zwei Teilen: a) der graphischen Darstellung der individuellen zeitlichen Entwicklung der Problembearbeitung (Zeitreihen) und b) der Beschreibung von Problemlösetypen. Diese Problemlösetypen beruhen auf der Klassifizierung der Zeitreihen und bezeichnen ein Leistungsniveau.

Erste Ergebnisse zur Überprüfung des Auswertungsmodells zeigen ermutigende Befunde: Die inhaltlich begründete, qualitative Klassifizierung der Zeitreihen (d.h. die Definition von Problemlösetypen) kann durch unterschiedliche formale Klassifizierungsmethoden bestätigt werden, und die Einschätzungen der Problemlöseleistung stimmen mit dem bestehenden Auswertungsansatz sowie der Beurteilung durch einen Experten überein. Letzteres weist darauf hin, dass die neue Auswertung die aufwändige und oft willkürliche Dateninterpretation und -integration durch den Diagnostiker ersetzen kann. Daneben werden auch Konsequenzen für die Grundlagenforschung diskutiert.

Computer-based problem-solving scenarios provide the opportunity to assess individual competences in dealing with complex, dynamic problems. This paper presents a new data analysis model, which claims to improve the practical significance of computer-based problem-solving scenarios in personnel selection by standardizing data analysis, data interpretation and data integration. The outcome is a process-oriented, micro-behaviour-based overall rating of the observed problem-solving activity. Furthermore, the overall rating is independent of the content of the scenario, and consists of two parts: a) a graphic display of the individual development of the problem-solving process in time (time-series), and b) the description of

problem-solving types. These types are based on the classification of the time-series and denote a level of problem-solving performance.

First results of testing the new data analysis model are encouraging: the qualitative classification of time-series (which means the definition of problem-solving types) can be confirmed by different formal classification methods, and the estimation of problem-solving performance corresponds with the existing data analysis approach as well as with the rating of an expert. The latter suggests that the new data analysis model may replace the laborious, time-consuming and often arbitrary data interpretation and data integration by the diagnostician. In addition, consequences for the fundamental research are discussed.

Danksagung

Allen, die in irgendeiner Weise zum Gelingen dieser Arbeit beigetragen haben, möchte ich an dieser Stelle herzlich danken.

Namentlich bedanken möchte ich mich an erster Stelle bei meinem Betreuer *Prof. Dr. René Hirsig* und meinem Arbeitskollegen *Dr. Thomas Wehrle*, die mich während meiner Zeit als Doktorandin weit über diese Arbeit hinaus begleitet und unterstützt haben. Die zahlreichen Diskussionen und Gespräche haben mein heutiges Verständnis einer Psychologischen Methodenlehre, der „Psychologie als Wissenschaft“ sowie meine generelle Lebensanschauung nachhaltig geprägt. *Prof. Dr. René Hirsig* danke ich zudem für die angenehme und wohlwollende Betreuung der Dissertation. Die vielen wertvollen inhaltlichen wie auch methodischen Anregungen halfen mir stets, eigene Überlegungen und Gedanken zu strukturieren und zu präzisieren. Dies sowie eine äusserst flexible Arbeitszeitgestaltung ermöglichten mir, die Dissertation in der dafür vorgesehenen Zeit von drei Jahren durchzuführen. *Dr. Thomas Wehrle* danke ich für sein grosses Interesse und aussergewöhnliches Engagement für diese Arbeit. Ohne die vielen Diskussionen und die häufig kritischen, aber konstruktiven Anregungen und Kommentare wäre diese Arbeit wohl in einem der vielen Tiefs, die bei einer Arbeit dieser Grösse unvermeidbar sind, stecken geblieben. Zudem liess sich ein Grossteil des methodischen Vorgehens nur dank des von ihm entwickelten Programms zur graphischen Darstellung der Zeitreihen und Berechnung der Ähnlichkeitsmasse realisieren.

Ein besonderer Dank gilt weiter *Dr. Astrid E. de With* für die Möglichkeit, das Problemlöseszenario ATLANTIS zu verwenden sowie für die Unterstützung bei diversen technischen und inhaltlichen Fragestellungen in diesem Zusammenhang. Die sich daraus ergebenden spannenden Diskussionen trugen ganz wesentlich zum Entstehen eines ersten Dissertationskonzepts sowie zur Entwicklung des neuen Auswertungsmodells (und insbesondere der Mikroverhaltensbeschreibung) bei.

Prof. Dr. Vinzenz Morger danke ich für die Übernahme des Ko-Referats und die genaue Durchsicht der vorliegenden Arbeit. Dank gebührt auch dem in dieser Arbeit

nicht namentlich erwähnten *Experten* für die interessante Zusammenarbeit und die aufwändige Beurteilung der beiden Probandengruppen.

Zu grossem Dank verpflichtet bin ich schliesslich meinem Mann *Alper Taner*, der alle Hoch und Tiefs der letzten Jahre mit mir geteilt hat und meine oft wechselnde emotionale Verfassung und innere Angespanntheit mit grenzenloser Geduld und grossem Verständnis getragen hat.

1 Einleitung

Viele Arbeiten zum „komplexen Problemlösen“ beginnen mit einer Ausführung über die zunehmende Komplexität der modernen Zeit und die damit verbundenen neuen Anforderungen an menschliches Denken und Handeln. An verschiedenen Alltagsbeispielen und Handlungsfeldern wird die *Bedeutung individueller Kompetenzen im Umgang mit komplexen Problemen und Systemen* illustriert. Gerade bei grösseren Arbeiten (wie Dissertationen) folgt nicht selten ein umfassender Rückblick auf die denkpsychologische Forschung, der zeigen soll, in welchem Ausmass diese Veränderung im menschlichen Leben auch die psychologische Forschung beeinflusst hat. Der Rückblick beinhaltet meist eine Erörterung unterschiedlicher Definitionsversuche, eine Gegenüberstellung von einfachen und komplexen Problemen, eine Beschreibung der verschiedenen Aufgabenstellungen und Aufgabenmerkmale, eine Übersicht über Theorien zum problemlösenden Denken, über die verschiedenen Paradigmen, Forschungsrichtungen und Befunde.

All dies wird in der vorliegenden Arbeit nicht erfolgen. Neben unzähligen solcher Einführungen existieren in der Zwischenzeit auch sehr gute Überblicksarbeiten, die sowohl die Geschichte wie auch den aktuellen Stand der gesamten Denk- und Problemlöseforschung sehr detailliert und umfassend darstellen. Erwähnenswert sind insbesondere der erst kürzlich erschienene Enzyklopädie-Band „Denken und Problemlösen“ von Joachim Funke (2006a) sowie weitere Arbeiten desselben (Funke, 2003; Funke & Frensch, 2005; Müller & Funke, 1995; oder in Englisch: Frensch & Funke, 1995).

Wenn von „komplexem Problemlösen“ die Rede ist, sind in der Regel Problemstellungen gemeint, die auf einem Computer simuliert werden und folgende Attribute aufweisen: Komplexität, Vernetztheit, Dynamik, Intransparenz und Polytelie. *Komplexität* meint die Komplexität der Anforderung, die Schwierigkeit der Problemstellung, die sich daraus ergibt, dass eine grosse Zahl verschiedener Aspekte (Variablen) für die Problemlösung relevant ist und berücksichtigt werden muss. *Vernetztheit* bezieht sich auf die Verknüpfung dieser unterschiedlicher Aspekte

beziehungsweise Variablen und verlangt das Berücksichtigen von Neben- und Fernwirkungen. *Dynamik* bezeichnet die Veränderung der Problemsituation über die Zeit hinweg, aufgrund der Handlungen der Versuchsperson oder auch einfach von selbst (Eigendynamik). Mit *Intransparenz* ist gemeint, dass die Problemsituation nicht vollständig durchschaubar ist, da nicht alle erforderlichen Informationen zu Beginn vorhanden sind (sowohl im Hinblick auf die beteiligten Variablen, wie auch im Hinblick auf die Zielvorgabe). *Polytelie* bedeutet Vielzieligkeit und beinhaltet, dass erfolgreiches Handeln in komplexen Situationen das gleichzeitige Verfolgen mehrerer Teilziele erfordert. Es reicht in der Regel nicht aus, nur ein Ziel im Auge zu behalten.

Seit den ersten Arbeiten und Überlegungen zu komplexen Problemstellungen in den 70er-Jahren, die im deutschsprachigen Raum vorwiegend initiiert waren durch Dietrich Dörner (vgl. z.B. 1976, 1989), werden diese Attribute komplexer Situationen mehr oder weniger unverändert in der Literatur aufgeführt (für eine ausführliche Beschreibung der einzelnen Attribute vgl. z.B. J. Funke, 2003, S. 126-135). Aus heutiger Sicht müssen jedoch nicht alle Attribute gegeben sein, damit ein Problem als komplex gilt. Wie J. Funke (2001) zeigen konnte, sind insbesondere die Merkmale der *Vernetztheit* und *Dynamik* charakteristisch für komplexe Problemstellungen. Für deren Realisierung ist ein Computer erforderlich, während die Merkmale *Intransparenz* und *Polytelie* auch bei einfachen Aufgabenstellungen vorliegen können. Auch ein Blick in die Literatur bestätigt, dass die verschiedenen computergestützten Problemlöseszenarien, die unter dem Begriff „komplexes Problemlösen“ aufgeführt werden, längst nicht alle diese Aspekte erfüllen.

An dieser Stelle bedarf es einer Begriffsklärung: Mit *computergestützten Problemlöseszenarien* (im Folgenden auch abgekürzt mit „Szenarien“) sind in dieser Arbeit komplexe, auf einem Computer implementierte Problemstellungen gemeint, die innerhalb der *komplexen Problemlöseforschung* entwickelt wurden. In der Literatur werden hierfür auch die Begriffe *(Problemlöse-) Simulationen*, *komplexe dynamische Systeme* oder *Mikrowelten* verwendet. Durch die denkpsychologische Verankerung grenzen sich Problemlöseszenarien klar von den ebenfalls unter den genannten Begriffen auffindbaren *computergestützten (Unternehmens-) Planspielen* beziehungs-

weise den *computer-assisted business/management games* ab. Auch wenn einige Planspiele gute Problemlöseszenarien darstellen und umgekehrt, so arbeiten die beiden Verfahrensgruppen doch mit unterschiedlichen Methoden und Zielsetzungen und werden nach unterschiedlichen Kriterien erstellt und vor allem auch ausgewertet (Kersting, 1999b). Planspiele entstammen der (betriebs-) wirtschaftlichen Tradition und sind im Unterschied zu den Problemlöseszenarien mehr als Lernspiele konzipiert, mit dem Ziel, bestimmte inhaltliche Fertigkeiten, Arbeitstechniken oder auch allgemeine Handlungskompetenzen zu vermitteln (Strauss, 2000). Dadurch beinhalten sie meist ein präziseres inhaltliches Realitätsmodell als Problemlöseszenarien. Planspiele werden vorwiegend zu Unterrichts- und Trainingszwecken und eher selten als diagnostisches Instrument eingesetzt. Während vereinzelt Problemlöseszenarien mit Planspielen gleichgesetzt oder erstere letzteren sogar untergeordnet werden (z.B. Geilhardt & Mühlbradt, 1995; Leutner, 1995), so wird mehrheitlich – und deshalb auch in dieser Arbeit – doch für eine Abgrenzung der beiden Konzepte plädiert.

Die in den vergangenen 30 Jahren entwickelten Problemlöseszenarien stammen aus den unterschiedlichsten Inhaltsbereichen wie Ökonomie, Ökologie, Biologie, Politik, Verkehr, Medizin oder Technik. Typische Rollen, mit denen Problemlöser¹ konfrontiert werden, sind beispielsweise die Führung eines Unternehmens oder einer Fabrik mit dem Ziel, möglichst viel Umsatz zu machen, die Verwaltung einer Stadt, die Bekämpfung eines Waldbrandes als Einsatzleiter der Feuerwehr, Kommandant sein einer Mondfähre oder die Unterstützung eines Stammes von Halbnomaden als Entwicklungshelfer. Neben der *semantischen Einbettung* unterscheiden sich die Szenarien hinsichtlich weiterer Kriterien, beispielsweise hinsichtlich

- formaler Merkmale (Anzahl Variablen, Art der Beziehung zwischen Variablen, formale Beschreibbarkeit der Systemstruktur etc.),
- Zielvorgabe (exakte, enge versus allgemeine, weite Ziele),
- Durchführung (mit oder ohne vermittelndem Versuchsleiter),
- Auswertung und Überprüfung dieser Auswertung.

¹ Hier wie in der gesamten Arbeit ist stets auch die weibliche Form gemeint.

J. Funke (2006b, S. 394-398) zeigt eine Auflistung der in der Literatur beschriebenen, knapp 100 (!) Szenarien, versehen mit mindestens einer Quellenangabe sowie Angaben über den Umfang des jeweiligen Szenarios, seine inhaltliche Einbettung und eventuelle „Familienzugehörigkeit“ (aus einigen Ursprungsszenarien wurden zahlreiche Varianten entwickelt).

Verschiedene taxonomische Systeme versuchen, etwas Ordnung in die Vielfalt an Szenarien zu bringen (für neuere Beispiele vgl. Gray, 2002; Jonassen, 2000; Wagener, 2001b). Bei vielen Szenarien fehlt allerdings eine ausführliche Dokumentation oder Referenz, wodurch die unterschiedlichen spezifischen Anforderungen, die an die Probanden gestellt werden, schwierig zu vergleichen sind. Dennoch lassen sich einige *zentrale Anforderungen* beschreiben, die komplexe Systeme im Allgemeinen an den Problemlöser stellen. Nach Dörner (1986) sind dies: *Informationsgewinnung und -integration, Zielausarbeitung und -balancierung, Massnahmenplanung und Entscheidung, sowie Selbstmanagement*. An diesem Anforderungsmodell hat sich grundsätzlich wenig verändert, wenn auch verschiedene Autoren einzelne Aspekte weiter ausdifferenziert oder anders umschrieben haben. Zusammenfassend werden folgende zentrale Anforderungen komplexer Probleme in der Literatur genannt (Schaub, 2006, S. 451):

- Zielbildung und Zielelaboration
- Absichtsauswahl und Schwerpunktbildung
- Informationssammlung
- Informationsintegration und Modellbildung
- Prognose und Extrapolation
- Planen und Entscheiden
- Umsetzung der Entscheidungen, Effekt- und Hintergrundkontrolle
- Kontrolle und Modifikation

Auch im Arbeitsalltag sind Menschen – je nach Beruf und Position unterschiedlich – mit solchen Anforderungen konfrontiert. Deshalb fanden die in der Forschung entwickelten Problemlöseszenarien von Anfang an auch im Bereich der *Personalarbeit* Anwendung, vorwiegend in der Personalauswahl (und im Speziellen in der Managementdiagnostik), aber auch in der Personalentwicklung und im Trainingsbereich (für einen Überblick vgl. Sammelband von Strauss & Kleinmann, 1995; und darin insbesondere der Beitrag von U. Funke, 1995). Daneben finden sich die Szenarien

aber auch in anderen Forschungs- und Anwendungskontexten wieder. Die Themen reichen von verschiedenen Aspekten des *Lernens* (De Jong & van Joolingen, 1998; Swaak, de Jong & van Joolingen, 2004), *webbasiertem Lernen* (Breuer, Molkenhuth & Tennyson, 2006) und *selbstgesteuertem Lernen* (Wirth, 2004), über *Emotionen* (Spering, Wagener & Funke, 2005) und *Persönlichkeitsmerkmale* (Schoppek & Putz-Osterloh, 2003) bis hin zur *Kultur vergleichenden Psychologie* (Strohschneider, 2001, 2006). Auch im Forschungsbereich zur *dynamischen Entscheidungsfindung* („Dynamic Decision Making“) erhalten die Szenarien eine zunehmende Bedeutung (Gonzalez, Vanyukov & Martin, 2005).

Auch wenn das zentrale Anwendungsfeld computergestützter Problemlöseszenarien im Bereich der *Personalauswahl* liegt, so haben die Szenarien innerhalb dieses Kontextes dennoch eine relativ geringe Bedeutung. Dies hat einerseits damit zu tun, dass die Szenarien im Vergleich zu anderen eignungsdiagnostischen Verfahren wenig ökonomisch sind. Die Anschaffung ist relativ teuer, zudem ist der Unterhalt wie bei allen computergestützten Verfahren aufwändiger als bei reinen Papier-Bleistift-Verfahren. Daneben erfordert die Anwendung der Szenarien viel Zeit und Erfahrung seitens der Diagnostiker. Die Ergebnisausgabe ist von Szenario zu Szenario sehr unterschiedlich, und oft werden auch innerhalb eines Szenarios eine Reihe unterschiedlicher Masse berechnet, wobei dem Diagnostiker oftmals keine oder ungenügende Angaben zur Dateninterpretation und -integration zur Verfügung stehen.

Andererseits gibt es nur wenige Untersuchungen über die testtheoretischen Gütekriterien von Problemlöseszenarien, die zudem sehr unterschiedliche Befunde liefern. Seit Beginn der 90er-Jahre wird deshalb die eignungsdiagnostische Brauchbarkeit der Szenarien verschiedentlich in Frage gestellt (J. Funke, 1998; Kersting, 1999a, 1999b, 2001; Kersting & Kluge, 2001; Kleinmann & Strauss, 1998; Süss, 1996, 1999).

Die vorliegende Arbeit stellt ein neues Auswertungsmodell vor, das sowohl die Anwendung der Szenarien in der diagnostischen Praxis erleichtern und verbessern soll, als auch eine mögliche Lösung für die heterogene Befundlage betreffend Reliabilität und Validität der Verfahren anbietet. Dieses, im theoretischen Teil der

Arbeit beschriebene Auswertungsmodell soll auf ein Szenario angewendet werden, das in der Personalauswahl eingesetzt wird. Dabei steht die Frage im Vordergrund, wie gut das neue Auswertungsmodell im Vergleich zur bestehenden, praxisbewährten Auswertung die Problemlöseleistung einzuschätzen vermag.

Wichtiger Bestandteil der neuen Auswertung ist die Standardisierung und Automatisierung der Dateninterpretation und Datenintegration. Resultat ist ein einfach verständliches, vom Inhalt des jeweiligen Szenarios unabhängiges Gesamturteil über die gezeigte Problemlöseleistung. Dieses Gesamturteil basiert auf einer Klassifizierung der individuellen Bearbeitungsprozesse der Probanden. Deshalb interessiert aus methodischer Sicht auch die Frage, wie gut das Gesamturteil durch unterschiedliche formale Klassifizierungsmethoden abgebildet werden kann. Aus der Anwenderperspektive interessiert, inwiefern das (programmierbare) Gesamturteil der neuen Auswertung die aufwändige Dateninterpretation und -integration durch den Diagnostiker ersetzen kann.

Die Arbeit ist folgendermassen aufgebaut: Kapitel 2 dient der Erläuterung der theoretischen Grundlagen und Überlegungen, die zur Entwicklung des neuen Auswertungsmodells geführt haben. Kapitel 3 beschreibt das methodische Vorgehen zur Anwendung und Überprüfung des neuen Auswertungsmodells. Die Ergebnisse dazu werden in den Kapiteln 4 und 5 dargestellt und diskutiert.

2 Theoretischer Hintergrund

Das neue Auswertungsmodell gründet in der vielleicht etwas gewagt anmutenden Behauptung, dass das Potential computergestützter Problemlöseszenarien zur Eignungsdiagnostik bisher nicht ausgeschöpft worden ist. Dieser Teil der Arbeit thematisiert diese Behauptung, indem aufgezeigt wird, worin die Defizite bestehender Auswertungsansätze liegen und mit welchen Schwierigkeiten die eignungsdiagnostische Anwendung der Szenarien (auch im Vergleich zu herkömmlichen eignungsdiagnostischen Verfahren) verbunden ist.

Kapitel 2.1 zeigt, welche Möglichkeiten zur Operationalisierung der Problemlöseleistung bisher genutzt wurden. Kapitel 2.2 führt in das Anwendungsfeld der Eignungsdiagnostik ein und erläutert die (zum Teil umstrittene) Bedeutung der Szenarien für diesen Bereich. Basierend auf den Erkenntnissen und Befunden der ersten beiden Kapitel wird in Kapitel 2.3 das neue Auswertungsmodell hergeleitet und vereinfacht vorgestellt. Das letzte Kapitel 2.4 erläutert die Fragestellungen und Hypothesen, mit denen das neue Auswertungsmodell überprüft werden soll.

2.1 Ansätze zur Erfassung komplexer Problemlöseleistungen

Wie in der Einleitung bereits dargestellt, gibt es nach gut 30 Jahren Forschung eine bald unüberschaubare Anzahl an Problemlöseszenarien, die sich hinsichtlich verschiedenster Kriterien unterscheiden. Angesichts dieser Vielfalt ist es wenig erstaunlich, dass auch die Meinungen darüber, wie komplexe Problemlöseleistungen zu messen sind, weit auseinander gehen. Es gibt nach wie vor keinen Konsens darüber, was mit den Szenarien gemessen werden soll und was nicht, und ob das, was gemessen wird, geeignete Indikatoren des Problemlöseerfolges sind, und was überhaupt unter „Erfolg“ beim Umgang mit komplexen Problemen zu verstehen ist. Die Indikatoren des Problemlöseerfolges variieren von Szenario zu Szenario, und oft wird auch innerhalb eines Szenarios eine Vielzahl unterschiedlicher und zum Teil widersprüchlicher Indizes berechnet. Welche Konsequenzen diese Vielfalt und Verschiedenheit vorhandener Problemlösegutemasse hat, zeigt das folgende Zitat

von Kersting (1999b):

Einige inkonsistente und unplausible empirische Ergebnisse im Kontext der Problemlöseforschung sind möglicherweise schlicht darin begründet, dass das Problem der Messung – welches jeder statistischen Analyse und inhaltlichen Interpretation vorausgeht – vernachlässigt wurde ... Die Vielzahl unterschiedlichster Problemlösegütemasse führt ausserdem dazu, dass unter dem Label „Problemlöseforschung“ de facto eine Vielzahl unterschiedlichster und unvergleichbarer Ansätze subsumiert werden ... Es ist alarmierend, dass aus ein und derselben Problembearbeitung – je nachdem, welcher Beurteilungsmassstab mit jeweils plausibler Begründung herangezogen wird – widersprüchliche Beurteilungen der Problemlösefähigkeit abgeleitet werden können. (S. 68-69)

Die bei der Bearbeitung anfallenden Ergebnisse werden üblicherweise drei Ebenen zugeordnet (U. Funke, 1995, S. 149; Höft & Funke, 2006, S. 151; Kersting, 1999b, S. 70-84): der Ebene der Steuerungsleistungen, der kognitiven Ebene und der Ebene der Verhaltensmasse und Strategien. Die Abgrenzung dieser drei Ebenen, insbesondere die Abgrenzung von Massen der kognitiven Ebene und Verhaltensmassen oder Strategien ist nicht immer eindeutig, wie die folgenden drei Abschnitte zeigen.

2.1.1 Erfassung der Steuerungsleistung

Masse für die Steuerungsleistung beziehen sich auf die unmittelbare *Erreichung des vom Szenario vorgegebenen Ziels*. Sehr häufig wird die Leistung über eine oder mehrere *zentrale Variablen* (so genannte Zielvariablen) gemessen, wobei zur Definition des Problemlöseerfolges entweder die Ausprägung dieser Variablen bei Spielende (z.B. das erreichte Firmenkapital) oder Abweichungsmasse (Vergleiche mit Optimallösungen, mit vorgegebenen Sollwertgrössen oder mit ungesteuerten „Nullläufen“) herangezogen werden. Nur vereinzelt wird auch der Prozesscharakter der Problembearbeitung betont, indem beispielsweise die durchschnittliche Abweichung von der Zielvariablen oder der zeitliche Verlauf bestimmter Variablen verwendet wird.

Liegen mehrere Zielvariablen vor, so können diese entweder einzeln betrachtet oder aggregiert werden. Bei *Aggregation* mehrerer Werte stellt sich allerdings das Problem der Gewichtung der einzelnen Komponenten und die Schwierigkeit der Entscheidung (und Begründung!) für eine bestimmte Berechnungsvariante. Eine weitere Schwierigkeit liegt darin, dass der *Wertebereich* der Variablen uneingeschränkt variie-

ren kann, was aufgrund von möglichen Ausreisserwerten und mangelndem Intervallskalenniveau zu Verzerrungen bei der Auswertung führen kann (da in vielen Arbeiten aufwändige parametrische Verfahren eingesetzt werden, z.B. lineare Strukturgleichungsmodelle, multiple Korrelationen etc.). Durch geeignete, theoretisch begründete Transformation der Variablen lassen sich solche Probleme jedoch (teilweise) umgehen.

Von solchen Indikatoren der Steuerungsleistung, die sich direkt an der Erreichung des vorgegebenen Ziels orientieren, unterscheidet Kersting (1999b, S. 71) so genannte *abgeleitete Masse*, die weniger eng an die Zielvorgabe gebunden sind. Beispiele hierfür sind Trendmasse wie die Anzahl an Monaten mit Aufwärtstrend im Flüssigkapital oder im Gesamtkapital. Nach Wagener (2001b, S. 79) handelt es sich bei diesen Trendmassen um „zyklusbezogene Transformationen“ von Zielvariablen. Die Zielvariable wird nach einer bestimmten Taktzahl transformiert und mit anderen Werten nach einem vorgegebenen, meist komplexen Schema verrechnet. Bei einer solchen Transformation kann gemäss Wagener (ebd.) Information verloren gehen, bei komplexeren zyklusbezogenen Transformationen können jedoch auch zusätzliche Informationen aus dem Verlauf der Problemlösung extrahiert oder Störeinflüsse gezielt unterdrückt werden. Insgesamt ist die Operationalisierung der Problemlöseleistung durch abgeleitete Masse aufgrund möglicher Verfälschbarkeit und mangelnder Transparenz für den Diagnostikanden und den Diagnostiker kritisch zu betrachten (Kersting, 1999b, S. 71-72).

Die Operationalisierung der Problemlöseleistung über Masse der Steuerungsleistung ist am meisten verbreitet. Masse der kognitiven Ebene sowie Verhaltensmasse und Strategien werden meistens *zusätzlich* zu den Steuerungsleistungen als Indikatoren für die Problemlöseleistung herangezogen. Vereinzelt erfolgt die Beurteilung der Problemlöseleistung auch nur anhand von Verhaltensmassen. Dies tritt vor allem dort auf, wo aufgrund der *Szenarienkonstruktion* (z.B. vorstrukturierter, mehr oder weniger fixer und damit von der Versuchsperson nicht beeinflussbarer Ereignisablauf) oder aufgrund *unspezifischer Zielvorgaben* (z.B. „Leiten Sie die Firma“) eine Bestimmung der Steuerungsleistung wenig sinnvoll oder nicht möglich ist. Beispiele hierfür

sind die Strategischen Management Simulationen (SMS-Szenarien; Streufert, Pogash & Piasecki, 1988) oder das Szenario STRATOS-00 (Hirsig & de With, 1991, 1992).

Befürworter der (ausschliesslichen oder zusätzlichen) Verwendung von Verhaltensmassen sehen den Vorteil ihres Vorgehens in der Möglichkeit, *Prozesskomponenten* zu diagnostizieren und damit das Zustandekommen von Ergebnissen differenziert erklären zu können. Tatsächlich kann man der reinen Verwendung von Massen der Steuerungsleistung den Vorwurf der „*Endprodukt*diagnostik“ machen: Registriert wird die Erreichung des vorgegebenen Ziels, beispielsweise das erreichte Gesamtvermögen (Szenario „DISKo“), der Prozentsatz vor dem Waldbrand geretteter Fläche (Szenario „Feuer“), die Bevölkerungszahl oder Grösse der Rinderherden (Szenario „Moro“) usw., eben der Ausprägungsgrad bestimmter Variablen oder Variablenkonstellationen bei Spielende (bei den genannten Szenarien werden allerdings neben diesen Steuerungsleistungen auch Verhaltensmasse berechnet). Nur selten wird für Steuerungsleistungen der zeitliche Verlauf der interessierenden Variablen dargestellt (vgl. z.B. Strohschneider, 1986), mehrheitlich erfolgt nur eine Aussage über das Endprodukt – analog zu psychometrischen Intelligenztests. Demgegenüber steht bei Verhaltensmassen und Massen der kognitiven Ebene der Prozess der Problembearbeitung im Vordergrund. Die folgenden beiden Kapitel erläutern, was darunter verstanden wird.

2.1.2 Erfassung kognitiver Aspekte der Problembearbeitung

Anhand der kognitiven Ebene sollen *Wissensstrukturen, mentale Modelle oder Denk- und Planungsprozesse* erfasst werden. Die Frage nach der (prädiktiven) Bedeutung von Wissen (über Variablen, Variablenrelationen, Variableneigenschaften oder über erfolgreiche Strategien und Heuristiken) für erfolgreiches Handeln in komplexen Situationen spielt dabei seit Beginn der komplexen Problemlöseforschung eine bedeutende Rolle. Dies hat damit zu tun, dass komplexes Problemlösen in einer sehr weit gefassten Definition als „Prozess des Erwerbs und der Anwendung von Wissen durch Handeln“ (Süss, 2001, S. 261) verstanden werden kann. Diese Definition erläu-

ternd, fährt Süss (ebd.) fort:

Auf der Grundlage der gegebenen Instruktionen und den vorgegebenen Zielen aktivieren die Probanden Vorwissen über den simulierten Bereich. Sie bilden Hypothesen über die Struktur des Systems und über die Auswirkungen möglicher Eingriffe ... Deklaratives Sachwissen muss in Handlungswissen und letztlich in konkrete Handlungsabsichten transformiert werden. Jeder ausgeführte Eingriff führt zu einer neuen Ausgangssituation, die mit den Erwartungen verglichen und bewertet werden muss. Es müssen induktive Schlussfolgerungen über das Systemverhalten abgeleitet werden, die Wissensbasis muss verfeinert, neue Hypothesen gebildet und durch konkrete Eingriffe geprüft werden (deduktives Denken). Diese Phase ist gekennzeichnet durch eine hohe Belastung des Arbeitsgedächtnisses, weil zahlreiche Informationen simultan gespeichert und verarbeitet werden müssen. (S. 261)

Nach Süss (1996, S. 67) sind *alle* theoretischen Ansätze zur Erklärung komplexen Problemlösens „Versuche, den Problemlöseprozess als Prozess des Wissenserwerbs und der Anwendung von Wissen zu beschreiben ... Die theoretischen Ansätze unterscheiden sich danach, welche Arten von Wissen angenommen werden, die Probanden bei der Systemsteuerung erwerben, welches Wissen sie zur Steuerung nutzen und wie vollständig dieses Wissen sein muss“. In Bezug auf die mit den Szenarien diagnostizierten *Wissensarten* (z.B. Sach- und Handlungswissen, deklaratives und prozedurales Wissen, bereichsspezifisches und bereichsübergreifendes, generelles Wissen) und die für diese Diagnose eingesetzten *Instrumente* gibt es grosse Unterschiede, was eine vergleichende Interpretation der Befunde erschwert (für aktuelle Überblicke vgl. J. Funke, 2003, Kapitel 5.2.4; Schoppek, 2002; Süss, 1996, 2001). Als unbestritten gilt heute, dass das *bereichsspezifische Vorwissen* die Problemlöseleistung signifikant beeinflusst (Schoppek & Putz-Osterloh, 2003, S. 165).

Für die Wissensdiagnose werden in der Regel *szenariospezifische Wissenstests* eingesetzt, die die Probanden vor, während oder nach der Problembearbeitung absolvieren müssen. Die Operationalisierung des Wissens bezieht sich dann meist – analog zu psychometrischen Intelligenztests – in irgendeiner Form auf die Anzahl richtig gelöster Aufgaben. Die Aufgaben sind je nach Art des zu erfassenden Wissens unterschiedlich gestaltet und beziehen sich auf das problemspezifische Vorwissen, das erworbene Wissen (Systemwissen) und die Wissensanwendung. So besteht ein *Wissenserwerbstest* aus reinen Wissensfragen, bei denen das in einer vorgängigen Systemexploration erworbene Wissen abgefragt wird (Bsp. „Was bewirkt....?“). Ein

Test zur Wissensanwendung besteht aus so genannten Steuerungsaufgaben, in denen das erworbene Wissen zur Erreichung des Aufgabenziels angewendet werden muss (Bsp. „Was müssen Sie machen, damit...?“), das heisst der Proband muss das System in einen gewünschten Zustand bringen. Deshalb wird das Mass für die Wissensanwendung auch mit der *Steuerungsleistung* gleichgesetzt (vgl. z.B. Kröner, 2001). Diese Art der Steuerungsleistung ist jedoch nur begrenzt mit den im vorangegangenen Kapitel beschriebenen Massen der Steuerungsleistung vergleichbar. Die Gemeinsamkeit besteht darin, dass sich beide auf die Erreichung des vorgegebenen Ziels beziehen. Bei der Steuerungsleistung als Wissensanwendung bezieht sich dieses Ziel jedoch nur auf die Steuerung des Systems *während bestimmter Aufgaben* (wobei jedes Aufgabenziel nur bestimmte Teilaspekte umfasst). Bei der Steuerungsleistung im eigentlichen Sinne (vgl. Kapitel 2.1.1) ist die gesamte Steuerung des Systems im Hinblick auf das generelle Spielziel gemeint.

An zwei Beispielen wird im Folgenden die getrennte Erfassung von Wissenserwerb und Wissensanwendung dargestellt.

Das von Kröner (2001) entwickelte System „Multiflux“ stellt eine fiktive Maschine dar, die aus je vier Reglern und Anzeigeinstrumenten besteht, die komplex miteinander verknüpft sind. Der Proband soll durch Explorieren die Verknüpfungen zwischen Reglern und Instrumenten herausfinden und anschliessend das System während einer bestimmten Zeit durch gezielte Eingriffe so steuern, dass sich die Instrumente möglichst exakt auf einem vorgegebenem Optimalwert einpendeln (Trennung von Explorations- und Steuerungsphase).

Kernstück der Auswertung von „Multiflux“ sind die folgenden drei Variablen (für weitere Variablen, die nur Varianten betreffen vgl. Kröner, 2001, S. 85-87):

1. Die *Explorationsstrategie*, operationalisiert als die *Differenz* aus a) dem Anteil von Takten der Explorationsphase, in denen gezielt drei von vier Reglern der Maschine auf Null gestellt wurden (dies entspricht einer gezielten experimentellen Bedingungsvariation, was positiv zu bewerten ist) und b) dem Anteil, bei dem alle vier Regler zugleich variiert wurden (was keiner effektiven Systemexploration dient und negativ zu bewerten ist).

2. Das *Systemwissen* (erworbenes Wissen), operationalisiert als Leistung im *Vorhersage-test* (Anzahl richtig gelöster Items; die Aufgaben beinhalten, für eine vorgegebene Situation mit eingestellten Reglern die Folgen der Reglerstellungen für die Instrumente vorherzusagen, Bsp. „Welchen Wert werden die Instrumente bei diesen Reglerstellungen im nächsten Takt annehmen?“).
3. Die *Wissensanwendung* (*Steuerungsleistung*), operationalisiert als Leistung im *Interventionstest* (wiederum Anzahl richtig gelöster Items; die Aufgaben beinhalten, für eine konkrete Situation diejenigen Reglerstellungen zu finden, die die Instrumente in den Optimalzustand versetzen, Bsp. „Wie müssen Sie die Regler einstellen, damit im nächsten Takt die angegebenen Zielwerte erreicht werden?“).

Ganz ähnlich sieht die Erfassung von Wissenserwerb und Wissensanwendung bei Wirth und Funke (2005) aus. Das für PISA-2000 entwickelte Szenario „Raumfahrtspiel“ besteht aus zwei zu steuernden Subsystemen, die analoge Strukturen aufweisen: einer Rakete, die zwischen vier Planeten hin und her fliegen kann und einem Planetenfahrzeug, das auf jedem der vier Planeten zu jeweils drei Diamanten und der Rakete fahren kann.

Wie bei Kröner (2001) wird der Wissenserwerb und die Wissensanwendung mit je zwei Tests gemessen (bestehend aus acht Interpolations- und acht prospektiven Aufgaben beim Wissenserwerbstest, sowie 22 Steuerungsaufgaben beim Wissensanwendungstest), wobei jeweils der prozentuale Anteil richtig gelöster Aufgaben die beiden *Erfolgsmasse* bilden. Daneben implementieren die Autoren auch so genannte *Regulationsmasse* für den Wissenserwerb und die Wissensanwendung. Diese geben als *Prozessmasse* Auskunft darüber, wie eine Person zum Erfolg gekommen ist.

Basierend auf der Arbeit von Wirth (2004) und Überlegungen zum selbstgesteuerten Lernen entwickeln die Autoren für den *Prozess des Wissenserwerbs* ein Mass, das auf der Basis von Verhaltensdaten die *selbstbestimmte Lernregulation* erfasst. Konkret wird für bestimmte Zeitintervalle erfasst, ob ein Lerner den Lernprozess stärker auf das *Identifizieren* von Informationen ausrichtet (neue Informationen entdecken) oder stärker auf das *Integrieren* von Informationen (identifizierte Informationen für späteres Abrufen „sichern“). Identifikation im Umgang mit finiten Automaten zeigt sich nach Wirth (2004, S. 86f.) durch Eingriffe, die zum ersten Mal ausgeführt werden, während Integration sich auf wiederholte Eingriffe bezieht. Anhand der vom System gespeicherten Rohdaten kann für jede Person für einen bestimmten Zeitraum das Verhältnis von identifizierenden und integrierenden Eingriffen berechnet werden

(beobachtbares Verhalten). Analog lässt sich angeben, welche Art von Eingriffsalternativen innerhalb eines Zeitraums häufiger zur Verfügung stand (erwartetes Verhalten). Als Mass für die selbstbestimmte Lernregulation werden diese beiden Quotienten zueinander in Beziehung gesetzt und logarithmisiert (Bezeichnung „ $\log_{(or)}$ “-Mass; o-r steht für „Odds Ratio“). Dadurch, dass der gesamte Lernprozess in mehrere Zeitintervalle eingeteilt wird, lässt sich der Verlauf dieses Masses mit Hilfe latenter Wachstumskurven modellieren (für eine ausführliche Beschreibung vgl. Wirth, 2004, S. 92f.). Gütemass für die Regulation des Wissenserwerbs (das Prozessmass) bildet der Wert des *Steigungsfaktors* des latenten Wachstumskurvenmodells für eine bestimmte Phase.

Für die Bewertung der *Regulationsgüte bei der Wissensanwendung* schlagen Wirth und Funke (2005) ein dem „ $\log_{(or)}$ “-Mass für den Wissenserwerb analoges Mass vor. In dieses geht ein, wie häufig ein Problemlöser einen Eingriff ausgeführt hat, der für die Zielerreichung optimal ist und wie häufig nicht optimale Eingriffe getätigt wurden, sowie wie viele optimale Eingriffe bekannt beziehungsweise nicht bekannt waren. Das Regulationsmass für die Wissensanwendung sowie Ergebnisse zu den verschiedenen Wissensmassen sollen an dieser Stelle nicht weiter vertieft werden. Der Ansatz ist für diese Arbeit deshalb interessant, weil er zeigt, dass sich testtheoretisch geeignete *verhaltensbasierte Prozessmasse* konstruieren lassen (zur Überprüfung der Gütekriterien vgl. Wirth, 2004; Wirth & Funke, 2005).

Die beiden beschriebenen Szenarien „Multiflux“ und „Raumfahrtspiel“ gehören einer Einteilung von J. Funke (2003, S. 146-162) zufolge zu den artifiziellen, formal orientierten Szenarien, die dieser den realitätsnahen, inhaltlich orientierten Szenarien gegenüber stellt. Die Konstruktion *artifizieller Szenarien* basiert auf einer formalen Struktur, konkret auf dem Ansatz linearer Strukturgleichungen (J. Funke, 1993) oder der Theorie finiter Automaten (Buchner, 1999; Buchner & Funke, 1993). Eine passende Semantik kann, muss aber nicht, im Nachhinein an die formale Struktur angepasst werden. Den Versuchspersonen wird die Künstlichkeit des zu explorierenden Systems oft explizit mitgeteilt.

Demgegenüber liegt den *realitätsnahen Szenarien* keine formal beschreibbare Systemstruktur zugrunde, im Vordergrund der Konstruktion steht der Inhalt, der nach-

zubildende Realitätsbereich. Erst im Nachhinein wird versucht, die wichtigsten Eigenschaften des Realitätsbereichs durch formale Modellierung (mehr oder weniger gut) abzubilden. Zu dieser Gruppe gehören beispielsweise die von der Arbeitsgruppe um Dörner erstellten Szenarien wie „Lohhausen“ (Dörner, Kreuzig, Reither & Stäudel, 1983), „Tailorshop“ und dessen Weiterentwicklungen zu „Textilfabrik“, „Schneiderwerkstatt“ etc. (für eine kurze Beschreibung der verschiedenen Tailorshop-Varianten vgl. U. Funke, 1995, S. 150-154) oder das Entwicklungshilfe-Szenario „Moro“ (z.B. Strohschneider, 1996, S. 75-77).

J. Funke (2003, S. 145) weist darauf hin, dass die Unterscheidung von realitätsnahen und artifiziellen Szenarien durchaus etwas willkürlich ist, da einerseits auch realitätsnahe Szenarien oft künstlich sind, und andererseits auch für viele formal orientierte Szenarien kompatible natürliche Systeme gefunden werden können. Im Zusammenhang mit der Wissensdiagnostik ist die Unterscheidung jedoch hilfreich, da nach J. Funke (ebd.) *nur artifizielle, formal orientierte Szenarien für eine umfassende Wissensdiagnostik geeignet sind*: Aufgrund der formal vollständig beschreibbaren Systemstruktur und der präzisen Formalisierung aller Anforderungen können *Wissenserwerb und Wissensanwendung* getrennt voneinander erfasst werden. Auch für inhaltlich orientierte Szenarien wurden Wissenstests entwickelt (für das häufig verwendete Szenario „Schneiderwerkstatt“ vgl. z.B. Kersting, 1999b; Süss, 1996). Diese erfassen jedoch nur das systemspezifische Vor- oder *erworbene* Wissen (in Form von Sach- und Handlungswissen unterschiedlicher Allgemeinheit) und machen keine Angaben zur Wissensanwendung. Letztere lässt sich bei inhaltlich orientierten Szenarien nur unzureichend erfassen, da die Systemstruktur häufig nicht vollständig beschreibbar ist und damit die vielfältigen Möglichkeiten zur Systemexploration durch den Probanden nur ungenau abgebildet werden können. Ausserdem lässt sich im Nachhinein bei vielen Untersuchungen mit inhaltlich orientierten Szenarien nicht entscheiden, ob es sich bei den diagnostizierten Wissensarten um Vorwissen oder um durch Systemexploration erworbenes Wissen handelt (vgl. z.B. Kersting, 2001).

2.1.3 Erfassung von Verhalten und Strategien

Auf der Ebene der Verhaltensmasse und Strategien interessieren *unterschiedlichste Angaben zum Umgang des Probanden mit der gestellten Aufgabe*, wie beispielsweise Angaben über das Frage- oder Entscheidungsverhalten, oder Angaben zum Informations- oder Selbstmanagement. Ausgezählt und bewertet werden die relativen oder absoluten Häufigkeiten von Verhaltenskategorien, ihre Verteilung, Veränderung über die Zeit, ihre Übergangswahrscheinlichkeiten usw. (Kersting, 1999b, S. 73). Grundlage dieser Berechnungen sind die im Rechnerprotokoll gespeicherten Rohdaten (Eingriffe/Aktionen des Probanden, Kennwerte bestimmter Variablen) und/oder Fremd- und Selbstbeobachtungen des Verhaltens während der Problembearbeitung.

Von der Vielfalt an Operationalisierungsversuchen ist die Ebene der Verhaltensmasse in besonderer Weise betroffen, wobei anzumerken ist, dass die Abgrenzung gegenüber Massen der kognitiven Ebene nicht immer eindeutig ist. So basiert beispielsweise die Erfassung der Regulationsmasse für den Wissenserwerb und die Wissensanwendung bei Wirth und Funke (2005; vgl. vorangegangenes Kapitel) auch auf Verhaltensdaten. Andererseits wird argumentiert, dass Strategien selbst Handlungswissen seien (Schoppek, 1996, S. 25). Auch für Süss (2001, S. 256) stellen Strategien Wissen dar: „Strategien, also zielgerichtete Denk- und Handlungsabläufe, sind in dieser Taxonomie dem Handlungswissen zuzuordnen, Pläne sind Strategien von hohem Allgemeingrad und Heuristiken vage Strategien von hohem Allgemeingrad“.

Im Folgenden sollen ein paar wenige Beispiele illustrieren, wie unterschiedlich die vorgeschlagenen Verhaltensmasse und Strategien sind. Einer Unterscheidung von Schmuck und Strohschneider (1995) zufolge werden zuerst ein paar Beispiele für „einfache Verhaltenseinheiten“ erläutert, anschliessend für „Konstellationen von Verhaltenseinheiten“, womit die Definition von Problemlösestilen, Problemlösetypen, Vorgehensweisen, Bearbeitungsmustern und Strategien gemeint ist. Es erklärt sich von selbst, dass die Vielfalt möglicher „einfacher Verhaltenseinheiten“ zu einer noch grösseren Vielfalt an möglichen „Konstellationen von Verhaltenseinheiten“ führt (vgl. auch Kersting, 1999b, S. 74).

Nach Schoppek (1991) soll die Leistung im Szenario „Feuer“ durch folgende Verhaltensmasse abgebildet werden: durch die Zahl der *Informationsabfragen* an den Computer, den Anteil aller *proaktiver Massnahmen* (= Planungsaktivitäten), den *Zeitbedarf* für die gesamte Bearbeitung, sowie durch die *Prioritätensetzung*, die *Operatorvielfalt* (betrifft Befehlskombinationen) und den Anteil *unsinniger Befehle*.

Für Dörner (1986) sind verschiedene *Indizes des Frage- und Entscheidungsverhaltens* geeignete Problemlösegütemasse. Beim Frageverhalten unterscheidet er Häufigkeit, Art und Ort der Frage, Hierarchien innerhalb von Fragereihen und Übergangshäufigkeiten von einem Fragetyp zum anderen. Beim Entscheidungsverhalten interessieren vor allem der Ort (der Inhaltsbezug) und die Dosierung der Entscheidung (die relative Stärke des Eingriffs). Daneben berechnet Dörner (ebd.) aus der Interaktion von Frage- und Entscheidungsverhalten weitere Indizes wie so genannte *Bündelungstendenzen* (berechnet durch die Häufigkeit des Überganges von Frage zu Entscheidung usw.) oder *Innovations- und Stabilitätsindizes* (die sich auf die Verschiedenheit bzw. Gleichheit von Entscheidungsspektren in aufeinander folgenden Takten beziehen, berechnet durch die Relation der zu einem Zeittakt jeweils neuen Fragen zur Anzahl Gesamtfragen).

Ähnlich geht Strohschneider (1986) vor und erfasst als relevante Verhaltensindizes beim Szenario „Moro“ die Gesamtzahl der Fragen sowie der Massnahmen, die Zahl unterschiedlicher Fragen und unterschiedlicher Massnahmen, den Bedarf an Hintergrundinformation sowie die Übergangswahrscheinlichkeiten Massnahme-Massnahme beziehungsweise Frage-Frage. Die Fragen und Massnahmen können wie bei Dörner nach Inhalt unterschieden werden, beispielsweise Fragen zur Bevölkerung, Landwirtschaft oder Umwelt oder Massnahmen in Bezug auf Soziales oder Handel etc. (Strohschneider, 1994). Daneben schlägt der Autor auch vor, verschiedene Verhaltensindizes und zentrale Kennwerte des Szenarios (Masse der Steuerungsleistung wie Bevölkerungszahl, Rinderzahl etc.) zu einer Art Gesamtbild zu integrieren und so *spezifische Vorgehensmuster und Bearbeitungstypen* zu beschreiben. Damit gelangt er zu Bezeichnungen wie *Stabilisator*, *Ausbeuter*, *Zerstörer*, *Verzettler* und *Zögerer* (Strohschneider, 1992, 1996).

Auch die Arbeitsgruppe um Dörner (Dörner et al., 2002, S. 263f.) definiert verschiedene Strategien, die die unterschiedlichen Vorgehensweisen der Versuchspersonen

bei der Bearbeitung des Szenarios „Insel“² beschreiben sollen: die *Basislager-Strategie*, die *Lageorientierung*, die *Action-Strategie*, die *Positions-Strategie*, die *Dagobert-Duck-Strategie*, die *Einkapselungs-Strategie*, die *Survival-Strategie* sowie einen *Wechsel der Strategien*. Für jede Strategie werden besonders kennzeichnende Verhaltensweisen genannt und anhand der Psi-Theorie (Dörner, 1999) das Entstehen der einzelnen Strategien erklärt. Verschiedenste Verhaltensmasse geben zudem Auskunft über den Erfolg einer Versuchsperson (Dörner et al., 2002, S. 261f.): die durchschnittliche Anzahl von *erfolgreichen sowie nicht erfolgreichen Bewegungen* von einem Ort zu einem anderen (wobei der Unterschied der beiden Zahlen ein Indiz für die Lernleistung ist), die Anzahl der verschiedenen von der Versuchsperson besuchten *Orte* (als Indiz für die Explorationstätigkeit), die Anzahl durchschnittlicher *Zusammenbrüche* des Roboters James (als Indiz für die Fürsorgepflicht), die Häufigkeit von *Objektannäherungen*, die Anzahl erfolgreicher sowie nicht erfolgreicher *Manipulationen* (wiederum als Indiz für die Lernleistung) und schliesslich die Anzahl gesammelter *Nukleotide*, sowie der *Hunger-, Durst und Schadenspegel* (als Bedürfnisindikatoren).

Diese wenigen Beispiele sollen genügen, um einen Eindruck von der Unterschiedlichkeit vorgeschlagener Verhaltenskategorien, Verhaltensmasse oder Strategieindikatoren zu vermitteln. Kersting (1999b, S. 73) weist darauf hin, dass die Definition und Bewertung dieser Masse nicht nur von Szenario zu Szenario variiert, sondern insbesondere auch von Autor zu Autor (für das gleiche Szenario) und nicht selten auch innerhalb der Arbeiten ein und desselben Autors! Daneben kritisiert Kersting (1999b, S. 75-84) drei weitere Punkte, die hier zusammenfassend dargestellt werden:

² Im noch wenig bekannten Szenario „Insel“ wird die Versuchsperson aufgefordert, auf einer kleinen Insel im Südpazifik mit Hilfe des ferngesteuerten Roboters „James“ nach so genannten „Nukleotiden“ zu suchen, die als der Brennstoff der Zukunft gelten. Zur Erkundung der Insel ist der Roboter mit verschiedenen Werkzeugen wie Greifarm, Sieb und Hammer ausgestattet, mit denen er Objekte manipulieren kann. Neben dem Sammeln von Nukleotiden muss sich die Versuchsperson auch um die Energieversorgung und die Funktionstüchtigkeit von „James“ kümmern, wobei eine Anzeige am Monitor über den aktuellen Durst-, Hunger- und Schadenspegel informiert. Ziel ist, möglichst viele Nukleotide zu sammeln und gleichzeitig „James“ am Leben zu erhalten.

1. *Fehlende theoretische Begründung und Einbettung, unklare Definitionen.* Es bleibt meistens unklar, inwiefern die ausgewählten Verhaltensweisen in der Praxis überhaupt vorkommen und inwiefern sie in der Praxis, oder zumindest im Szenario, überhaupt zu Problemlöseerfolg führen. Besonders unbefriedigend ist der Begriff *Strategien* und die verschiedentlich vorgenommene, oft willkürliche Definition von *Problemlösetypen*: Sowohl Bereichsunabhängigkeit wie auch Bereichsabhängigkeit wird in verschiedenen Arbeiten als charakteristisches Merkmal von Strategien aufgeführt, und die Abgrenzung gegenüber Begriffen wie Problemlösestile, -typen oder auch Wissen über Strategien ist unklar (s. dazu auch Bemerkung weiter oben).
2. *Beliebigkeit der Ableitung und Bewertung von Verhaltensweisen, mangelnde oder fehlende Standardisierung und Auswertungsobjektivität.* Die Zuordnung einzelner Daten zu den Datenkategorien erscheint oft willkürlich, eine ausführliche oder zumindest einigermaßen nachvollziehbare Dokumentation fehlt meistens, so dass die Ergebnisse nicht repliziert werden können. Zudem existieren auch für die Auswertung und Interpretation der (oft widersprüchlichen) Indikatoren meistens keine Regeln.
3. *Unabhängigkeit von Verhaltensmassen und Steuerungsleistungen.* In den meisten Studien kann kein substantieller Zusammenhang zwischen den Verhaltensmassen und dem Steuerungserfolg beziehungsweise der instruktionsgemässen Zielerreichung nachgewiesen werden.

Während die ersten beiden Punkte durchaus ihre Berechtigung haben, muss der letzte (vom Autor als besonders gravierend dargestellte) Punkt relativiert werden. Nach U. Funke (1995) sprechen verschiedene Faktoren dagegen, eine allzu hohe Übereinstimmung zwischen den verschiedenen Auswertungsebenen zu fordern. Szenarien verfügen nicht über eine einzige richtige Lösung, „sondern Erfolge sind meist auf verschiedenen Wegen und damit auch in verschiedenen Konstellationen einzelner Ergebnismasse erzielbar“ (S. 182). Zudem können die Wirkungspfade vom Verhalten zur Steuerungsleistung unterschiedlich direkt sein – beispielsweise wird die Steuerungsleistung vom Entscheidungsverhalten eher direkt, vom Informationsverhalten eher indirekt bestimmt – weshalb die Zusammenhänge *unterschiedlich eng* ausfallen können.

Dies bestätigt auch die Studie von Wagener und Wittmann (2002). Das von den Autoren verwendete Szenario „FSYS“ (Wagener, 2001b) beruht auf dem Anforderungsmodell von Dörner (1986) und beinhaltet eine Skala für das *Gesamtergebnis* (die Steuerungsleistung), je vier Subskalen für die Skala *Massnahmengüte* und die Skala *Informationsmanagement* sowie drei Subskalen für die Skala *Selbstmanagement*. Anhand eines Pfadmodells können die Autoren nachweisen, dass das Gesamtergebnis der Problemlöseleistung in erster Linie von der Qualität der tatsächlich getroffene-

nen Massnahmen abhängig ist, während Informations- und Selbstmanagement indirekt über die Massnahmenqualität das Gesamtergebnis beeinflussen. Die Höhe des Zusammenhangs zwischen den Verhaltensskalen der Massnahmengüte und der Steuerungsleistung (Gesamtvermögen) liegt für drei der vier Subskalen zwischen $r = .41$ bis $r = .63$, was durchaus als zufrieden stellend betrachtet werden darf, geht man von Validitätsanforderungen nach Weise (1975, S. 219; zit. nach Bortz & Döring, 2002, S. 201) oder der Effektstärkenklassifikation nach Cohen (1992, S. 157) aus. Auch in anderen Studien zeigen sich Zusammenhänge in vergleichbarer Höhe zwischen *einzelnen* Verhaltensmassen und Massen der Steuerungsleistung (z.B. Jansson, 1994; Putz-Osterloh, 1993; Schaub, 1990).

Die verschiedenen Masse erfassen gemäss U. Funke (1995, S. 182) unterschiedliche Aspekte der Leistung im Szenario, die *alle* berücksichtigt werden sollten. Die ausschliessliche Erfassung der Steuerungsleistung ohne Berücksichtigung von Verhaltensmassen ist gerade auch für die diagnostische Praxis wenig interessant. U. Funke (ebd.) stellt fest, dass manche Probanden rein ergebnisorientiert vorgehen und eine Massnahme, sobald sie damit einmal Erfolg gehabt haben, immer wieder wiederholen. Damit können sie womöglich günstige Gesamtergebnisse erzielen, während die Verhaltensergebnisse weniger gut ausfallen werden. Für diagnostische Zwecke sind differenzierte Aussagen zu verschiedenen Aspekten der Problembearbeitung hilfreich. Diesem Einsatzgebiet computergestützter Problemlöseszenarien widmet sich das nächste Kapitel.

2.2 Problemlöseszenarien als eignungsdiagnostische Instrumente

Nur kurze Zeit, nachdem sich Problemlöseszenarien in der psychologischen Forschung etabliert hatten, wurden die neuen Instrumente auch in der eignungsdiagnostischen Praxis und im Trainingsbereich eingesetzt (für einen Überblick vgl. U. Funke, 1995). Die anfängliche Begeisterung für die neuen Instrumente war enorm. Allerdings lässt sich feststellen, dass nur wenige Szenarien seitens der Wissenschaft speziell für diagnostische Zwecke entwickelt wurden. Die erhöhte Nachfrage wurde

vorwiegend durch Verfahren befriedigt, die entweder aus der Grundlagenforschung übernommen oder von kleineren kommerziellen Anbietern in eigener Regie entwickelt wurden (Wagener & Wittmann, 2002, S. 80). Eine Prüfung der testtheoretischen Gütekriterien blieb in der Mehrheit der Fälle aus. Vor allem seitens der Wissenschaft wurde deshalb die eignungsdiagnostische Verwendung der Szenarien in der Folgezeit kritisch diskutiert (J. Funke, 1998; Kersting, 1999a, 1999b, 2001; Kersting & Kluge, 2001; Kleinmann & Strauss, 1998; Süß, 1996, 1999).

In dieser Arbeit steht die eignungsdiagnostische Anwendung der Szenarien im Rahmen der *Personalauswahl*, und dabei speziell der *Managementdiagnostik*, im Vordergrund. Kapitel 2.2.1 dient einer Einführung in dieses Anwendungsfeld, indem die wichtigsten Konzepte der Berufseignungsdiagnostik kurz erörtert werden. Da die Problemlöseszenarien – wie auch das in dieser Arbeit verwendete Szenario – meist im Rahmen von so genannten Assessment Centern angewendet werden (Kleinmann & Strauss, 1998, S. 99), wird dieser Ansatz etwas ausführlicher dargestellt. Anschließend folgt eine Erläuterung der Argumente, die für einen Einsatz der Szenarien in der Personalauswahl sprechen (Kapitel 2.2.2). Die Überprüfung der eignungsdiagnostischen Qualitätsanforderungen der Szenarien steht im Mittelpunkt des letzten Kapitels (Kapitel 2.2.3). Neben den wichtigsten Befunden aus entsprechenden Untersuchungen und möglichen Erklärungen für diese Befunde wird auch die Frage diskutiert, inwiefern Problemlöseszenarien gegenüber herkömmlichen eignungsdiagnostischen Verfahren einen Mehrwert aufweisen.

2.2.1 Begriffsdefinition: Personalauswahl und Managementdiagnostik

Grundlage der Personalauswahl ist die *Berufseignungsdiagnostik*, die von Schuler und Höft (2004a) folgendermassen definiert wird:

Berufseignungsdiagnostik bezeichnet die Methodologie der Entwicklung, Prüfung und Anwendung psychologischer Verfahren zum Zwecke eignungsbezogener Erfolgsprognosen und Entscheidungshilfen im beruflichen Kontext ... Praktische Einsatzbereiche sind alle Arten berufsbezogener Einschätzung, Beratung, Selektion und Zuordnung mit dem Ziel der individuellen Berufs-, Organisations- und Arbeitsplatzwahl, der institutionellen Berufsberatung sowie der Auswahl und Entwicklung von Mitarbeitern in Wirtschafts- und Verwaltungsorganisationen. (S. 440)

Drei in wechselseitiger Beziehung stehende Kernbereiche bestimmen die eignungsdiagnostische Arbeit: die aus den beruflichen Tätigkeiten (anhand von Arbeits- und Anforderungsanalysen) abgeleiteten *Anforderungen*, die gewählten *eignungsdiagnostischen Verfahren* sowie die zu prognostizierenden *Kriterien des Berufserfolges* (Schuler & Höft, 2004a, S. 443; basierend auf Schuler & Funke, 1993). Durch Arbeits- und Anforderungsanalysen werden Informationen zur Berufstätigkeit gesammelt, die anschliessend zur Konstruktion unterschiedlicher eignungsdiagnostischer Verfahren und leistungsbezogener Beurteilungskriterien dienen.

Die eignungsdiagnostischen Instrumente lassen sich in drei Verfahrensklassen aufteilen (z.B. Schuler & Höft, 2006, S. 103): Mit dem *Eigenschaftsansatz* sollen Merkmale erfasst werden, die als relativ stabil angenommen werden, typische Messverfahren sind psychometrische Tests (Persönlichkeits-, Intelligenztests). Mit dem *Simulationsansatz* soll die Leistungsfähigkeit in konkreten beruflichen Anforderungen erfasst werden, typische Beispiele sind Arbeitsproben, Postkorbübungen, Rollenspiele, Gruppendiskussionen oder Präsentationen. Der *biographische Ansatz* schliesslich beinhaltet Fragen zur Biographie des Probanden, entweder in Form von Fragebogen oder mündlichem Interview.

Für die Unterscheidung von eigenschaftsorientierten und simulationsorientierten Verfahren eignet sich die Auflistung von U. Funke (1993, S. 111, 1995, S. 174; vgl. auch Höft & Funke, 2006, S. 146), die in Tabelle 1 wiedergegeben ist. Hinsichtlich der darin aufgeführten Merkmale sind computergestützte Problemlöseszenarien nach U. Funke (ebd.) eindeutig dem simulationsorientierten Ansatz zuzuordnen. Einzig hinsichtlich des Bezugs zum Arbeitsplatz ist die Zuordnung weniger klar, da die Realitätsnähe vieler Szenarien nicht sehr gross ist. Im Vergleich zu klassischen Intelligenztests dürfte die Realitätsnähe aber dennoch grösser sein.

Eignungsdiagnostische Verfahren finden in der Personalauswahl selten als Einzelverfahren Anwendung. In der Regel werden sie kombiniert im so genannten Assessment-Center-Ansatz eingesetzt. Unter einem *Assessment Center (AC)* versteht man „eine systematische und für den jeweiligen Einsatz massgeschneiderte Verfahrenstechnik, in der der „Multi“-Gedanke eine herausragende Bedeutung spielt: Es werden simultan mehrere Teilnehmer von mehreren Beobachtern in

mehreren Verfahren hinsichtlich mehrerer definierter Anforderungen beurteilt“ (Höft & Funke, 2006, S. 162). Ziel eines ACs ist entweder die Einschätzung der aktuellen Kompetenzen (Personalauswahl-AC) oder die Prognose der künftigen beruflichen Entwicklung der beurteilten Person (Entwicklungs-AC).

Tabelle 1: Vergleich von eigenschafts- und simulationsorientierten Verfahrenskonzepten (nach U. Funke, 1993).

	Eignungsdiagnostische Verfahrenskonzepte	
	eigenschaftsorientiert (Test)	simulationsorientiert (Arbeitsprobe)
Bezug zum Arbeitsplatz	geringer (eigenschaftsbezogen)	höher (tätigkeitsbezogen)
Anforderungsbereich	eigenschaftshomogen	eigenschaftsheterogen
Ablaufstruktur	Einzelitems	Gesamtaufgabe (ganzheitlich)
Rückmeldung, Kontrolle	gering	Ablauf-Feedback, Ergebnis-Feedback
Verlaufscharakter	statisch	dynamisch (reaktionsabhängige Aufgabenentwicklung)
Typische Beispiele	Intelligenztest, Persönlichkeitsverfahren	Gruppendiskussion, Postkorb, diagnostisches Rollenspiel

Ein AC kann ein- oder mehrtägig sein und unterschiedlich viele, mehr oder weniger realitätsnah gestaltete Übungen umfassen (für ausführliche Überblicke über die AC-Methode vgl. Kleinmann, 2003; Obermann, 2002; Sarges, 2001b). Das theoretische Konzept des AC-Ansatzes geht von einer starken *Verhaltensorientierung* der Eignungsdiagnose aus (Schuler & Höft, 2004a, S. 474). Deshalb wird auch immer wieder die (wenn möglich) alleinige Verwendung simulationsorientierter Verfahren propagiert (vgl. dazu die vom Arbeitskreis Assessment Center herausgegebenen Standards der AC-Methode, 1992, 2004). In der Praxis kommen jedoch trotzdem neben simulationsorientierten Verfahren auch eigenschaftsorientierte sowie biographische und Interviewverfahren zur Anwendung. Die Vernachlässigung psychologischer Testverfahren im AC-Ansatz wird zudem stark kritisiert (z.B. Hossiep, 2001).

Eine besondere Rolle bei der Durchführung eines ACs kommt den *Beobachtern* zu (Höft & Funke, 2006, S. 167-172): Die Beobachter haben die Aufgabe, nach jeder Übung die Teilnehmerleistung hinsichtlich der vorgesehenen Anforderungsdimensionen zu beurteilen (meist in Form numerischer Urteile, teilweise auch in Form verfahrensspezifischer Verbalurteile). In der klassischen AC-Variante wird dabei die Beobachter-Teilnehmer-Zuteilung (im Verhältnis 1:2, d.h. sechs Beobachter auf zwölf Teilnehmer) über ein vorgängig festgelegtes Rotationsverfahren nach jeder Übung geändert, so dass am Ende des ACs womöglich jeder Beobachter jeden Teilnehmer in mindestens einer Übung beurteilt haben sollte. Am Ende des ACs wird für jeden Teilnehmer ein *Gesamturteil* erstellt. Am häufigsten kommen dabei Konsensdiskussionen zum Einsatz. Alternativ werden auch (von einer Minderheit) statistische Hilfsmittel zur Urteilsbildung eingesetzt.

Bei der Verwendung von computergestützten Problemlöseszenarien im AC unterscheiden Strauss und Kleinmann (2001, S. 70) drei Einsatzgebiete, in denen den Beobachtern unterschiedliche Aufgaben zukommen: Neben den selteneren Einzel-ACs (Variante 1) kommen die Verfahren vor allem in Gruppen-ACs zur Anwendung, wobei die Bearbeitung des Szenarios hierbei entweder von allen Testpersonen *einzel*n (jeder für sich wie bei den Einzel-ACs, Variante 2) oder von mehreren Testpersonen *zusammen* (als Gruppe, Variante 3) erfolgen kann. Bei den Varianten 1 und 2 (ein Arbeitsplatz, eine Person) sind in der Regel während der Szenariendarstellung keine Beurteiler anwesend und Grundlage der abschliessenden Beurteilung bilden die vom Programm automatisch erstellten Indizes (über die Güte der Problemlösung oder Angaben zum Verhalten). Bei der Variante 3 (ein Arbeitsplatz, mehrere Testpersonen) sind Beurteiler anwesend, die das Problemlöseverhalten in der Gruppe nach bestimmten Verhaltenskriterien beurteilen.

Die am weitesten verbreitete Funktion der AC-Methode ist die Auswahl und Förderung von *Führungskräften* (Schuler & Höft, 2004b, S. 308). Dabei dienen ACs „der Einschätzung von derzeitigen Management- (bzw. Führungs-) Kompetenzen (= Diagnose) bei Problemen der Entwicklung und von zukunftsbezogenen Management- (bzw. Führungs-) Potential (= Prognose) bei Problemen der Auswahl von Kandidaten“ (Sarges, 2001a, S. VII). Innerhalb der *Managementdiagnostik* gelten ACs –

aus der Perspektive der Wissenschaft wie auch der Praxis – als die validesten eignungsdiagnostischen Instrumente (Sarges, ebd.). Überblicke über weitere in der Managementdiagnostik eingesetzte Verfahren finden sich bei Sarges (2000) sowie in verkürzter Version bei Wottawa und Hossiep (1997, Kapitel 5).

Da gerade in ACs zur Auswahl und Förderung von Führungskräften und Managern häufig computergestützte Szenarien zum Einsatz kommen, erläutert das folgende Kapitel, weshalb die Szenarien hierfür besonders interessant sind. Daneben werden auch Gründe aufgeführt, die ganz allgemein für die eignungsdiagnostische Anwendung der Szenarien sprechen.

2.2.2 Vorteile und Nutzen der Szenarien

Eines der am häufigsten vorgebrachten Argumente für den Einsatz computergestützter Szenarien in der Managementdiagnostik gründet in ihrem Anspruch, die gleichen (oder zumindest ähnliche) *Anforderungen* an die Testperson zu stellen wie reale Führungs- und Managementaufgaben. So gilt der Umgang mit Komplexität, Intransparenz, Dynamik und Vernetztheit als wesentlicher Bestandteil des Berufsalltags von vielen Führungskräften und Managern. Die vielfältigen Anforderungen, mit denen diese Personen konfrontiert werden, gleichen den von Problemlöseszenarien gestellten Anforderungen, wie beispielsweise Informationssammlung, Informationsintegration, Zielausarbeitung, Zielbalancierung, Massnahmenplanung, Entscheiden, Umsetzen, Kontrolle oder auch Selbstmanagement (vgl. Kapitel 1). Solche Anforderungen werden in Stellenausschreibungen häufig in Etiketten wie „Fähigkeit zum vernetzten Handeln/Denken“, „Fähigkeit zum systemischen Denken“, „Strategische Kompetenzen“ oder „Hohe heuristische Kompetenz“ zum Ausdruck gebracht.

Wenn auch die Anforderungskorrespondenz von Szenarien und Berufsaufgaben (Kriterium der Inhaltsvalidität) verschiedentlich kritisiert wird (z.B. Kersting, 1999b, 2001), so muss doch zumindest eingestanden werden, dass herkömmliche eignungsdiagnostische Verfahren keine vergleichbaren Aussagen über den Umgang mit

Komplexität und Dynamik ermöglichen. Eigenschaftsorientierte (psychometrische) Verfahren wie Intelligenz- und Persönlichkeitstests sind grundsätzlich *statusorientiert* und beschreiben Endprodukte. Mögliche diagnostische Aussagen beziehen sich lediglich auf die mit dem jeweiligen Test operationalisierte spezifische Eigenschaft (z.B. allgemeine Intelligenz, Merkfähigkeit, Verarbeitungskapazität). Simulationsorientierte Verfahren wie Postkorbübungen, Gruppendiskussionen oder Rollenspiele sind zwar *prozessorientiert* und ermöglichen Aussagen über unterschiedlichste *arbeitsbezogene Fertigkeiten*, ihre Auswertung und Interpretation basiert aber fast ausschliesslich auf *Beobachterurteilen*. Demgegenüber bieten Szenarien die Möglichkeit, das während der Problembearbeitung geäusserte Verhalten *automatisch* aufzuzeichnen und analog zu computergestützten psychometrischen Verfahren die Beurteilung der Problemlöseleistung *automatisch* in Form von (je nach Szenario unterschiedlichen) Leistungswerten auszugeben. Werden als Leistungswerte nicht nur reine Erfolgsangaben in Form von Steuerungsleistungen verwendet, sondern auch Aspekte des Verhaltens berücksichtigt, so steht eine grosse Menge *zusätzlicher Informationen* zur Verfügung. Eine vergleichbare Informationsquelle lässt sich mit psychometrischen Verfahren nicht generieren. Zudem sind Verhaltensdaten weniger anfällig für verschiedene Methoden der Testverfälschung (wie soziale Erwünschtheit, Antworttendenzen etc.), die bei psychometrischen Verfahren häufig beobachtet werden (Wagener, 2001b, S. 141).

Für die diagnostische Arbeit sind solche zusätzlichen Informationsquellen aus verschiedenen Gründen relevant (U. Funke, 1995, S. 150). Verhaltensdaten können mögliche Erklärungen für das Zustandekommen der Ergebnisse (in der Regel der Steuerungsleistungen) liefern und so auf Ansatzpunkte für Verhaltensänderungen hinweisen. Daraus lassen sich Ideen für ein detailliertes und konstruktives diagnostisches Abschlussfeedback oder für Personalentwicklungsmassnahmen ableiten.

Wagener und Wittmann (2002, S. 91-92) erläutern den zusätzlichen Ertrag von Verhaltensskalen für die Personalentwicklung an zwei Beispielen. Zwei Testpersonen mit dem gleichen Wert für das Gesamtergebnis (Steuerungsleistung) zeigen unterschiedliche Verhaltensprofile, das heisst unterschiedliche Ausprägungen in verschiedenen Verhaltensskalen. Während Person A die Prioritäten deutlich besser

setzt und mehr Aktivitäten zur Informationsbeschaffung zeigt, setzt Person B ihre Massnahmen effektiver um. Aus solchen Stärken- und Schwächenprofilen können nun gezielte Trainingsmassnahmen abgeleitet werden, beispielsweise könnte der Person B die konsequente Nutzung von Informationsquellen vermittelt werden und Person A müsste auf der taktischen Ebene geschult werden.

Ein weiterer Vorteil von Verhaltensskalen liegt in der Anpassung des Testmaterials an die Fragestellung: Je nach Anforderungen, die eine zu besetzende Position stellt, können die passenden Verhaltensskalen ausgewählt und entsprechend gewichtet werden (Wagener, 2001a, S. 70).

Ein gerne aufgeführtes Argument für den diagnostischen Einsatz von Problemlöseszenarien betrifft deren *soziale Akzeptanz*. Verschiedene Untersuchungen weisen darauf hin, dass der Einsatz computergestützter Szenarien sowohl auf Seiten der Kandidaten wie auch der Unternehmen als positiv eingeschätzt wird (Hartung & Schneider, 1995; Schöning, 1995; Schreiber, 1995). Im Vergleich zu Intelligenztests werden Problemlöseszenarien – der Untersuchung von Kersting (1998) zufolge – bezüglich „Anforderungs- und Realitätsnähe“ und bezüglich „positivem Erleben“ (Spass beim Bearbeiten) von den Teilnehmern besser beurteilt. Intelligenztests übertreffen die Szenarien jedoch in Bezug auf die „Kontrollierbarkeit“, wobei den Intelligenztests eine höhere Qualität der Messung zugesprochen wird. Demnach verfügen Problemlöseszenarien nicht, wie verschiedentlich angenommen, über eine höhere Akzeptanz als Intelligenztests, sondern jedes der beiden Verfahren verfügt über *spezifische (Akzeptanz-) Vor- und Nachteile*. Zudem sind die Akzeptanzeinschätzungen gemäss Kersting (ebd.) abhängig von verschiedenen Personenfaktoren (wie dem Alter, der eigenen Leistung im jeweiligen Verfahren oder der Einstellung gegenüber Computern). Nur die Einschätzungen des allgemeinen positiven Erlebens erwiesen sich als unabhängig von solchen Einflüssen. Entsprechend wird auch die *hohe Motivation*, die die Szenarien bei den Testpersonen auslösen, als weiteres Argument für den diagnostischen Einsatz der Szenarien aufgeführt (J. Funke, 1998, S. 93; Hornke & Kersting, 2006, S. 154).

Nach der Erläuterung einiger Vorteile, mit denen der Einsatz computergestützter Problemlöseszenarien in der Eignungsdiagnostik verbunden ist, zeigt das nächste Kapitel, worin die Probleme und Schwachstellen dieser Anwendung liegen.

2.2.3 Qualitätsanforderungen und ihre Überprüfung

Werden Szenarien für eignungsdiagnostische Zwecke eingesetzt, so sind die üblichen Anforderungen an sie zu stellen wie an andere Testinstrumente auch. Nach Hasselmann (1995, S. 239) müssen folgende Qualitätsanforderungen erfüllt sein:

- definierte und relevante Anforderungen
- ausführliche Dokumentation des Systemverhaltens
- standardisierte Instruktion
- leichte, ergonomische Bedienbarkeit
- Objektivität der Durchführung und Auswertung
- stabile und reliable Gütemasse
- standardisierte und normierte Auswertung
- empirisch fundierte Interpretationshinweise
- empirischer Nachweis der (prädiktiven) Validität

Die ersten vier Punkte beziehen sich gemäss Hasselmann (ebd.) dabei insbesondere auf den Einsatz der Szenarien in Seminarien, Trainings oder als Gruppenübung im AC, die letzten vier Punkte auf den Einsatz in der Eignungsdiagnostik. Bereits die ersten vier Anforderungen werden von den meisten Szenarien jedoch nicht erfüllt, und Studien zur Überprüfung der übrigen Anforderungen (d.h. der testtheoretischen Gütekriterien) sind vereinzelt positiv, aber leider selten (Strauss & Kleinmann, 2001). J. Funke (1999, S. 196) stellt diesbezüglich fest: „Der fast unüberschaubare Markt von Simulationsszenarien, die in eignungsdiagnostischen Kontexten zur Anwendung gelangen, ist voll von unüberprüften und vermutlich untauglichen Exemplaren“.

Eine viel ausführlichere und längere Liste an Voraussetzungen für den eignungsdiagnostischen Einsatz von Szenarien findet sich bei Kersting (1999b, S. 260) sowie zusammengefasst bei Kersting und Kluge (2001, S. 81). Für die Konstruktion neuer Szenarien und die Überprüfung bestehender Szenarien sind solche Anforderungslisten unabdingbar. In Anbetracht dessen, dass für die grosse Mehrheit der heute in

der Praxis verwendeten Szenarien die verschiedenen Anforderungen nicht einmal ansatzweise erfüllt sind, werden die Anforderungslisten hier nicht weiter vertieft.

Im Folgenden werden die Qualitätsmängel der Szenarien aufgegriffen, die in der Literatur am häufigsten diskutiert werden: das Prüfen der *Reliabilität* und *Validität* der Instrumente. Da die verschiedenen (wenigen) Studien zur Überprüfung dieser beider Kriterien bereits mehrfach zusammengefasst und diskutiert wurden (z.B. J. Funke, 2006b; U. Funke, 1995; Kersting, 1999b; Kleinmann & Strauss, 1998; Naef, 2001b; Strauss, 2000; Süss, 1996), wird hier auf eine Wiederholung der einzelnen Studien verzichtet. Insgesamt sind die Befunde sehr heterogen. Die Methoden zur Reliabilitäts- und Validitätsbestimmung sind zum Teil sehr unterschiedlich und den Studien haften nicht selten erhebliche methodische Mängel an (zu kleine Stichproben, methodisch ungenügende Auswertung).

Zuerst folgt ein Überblick über die wichtigsten Befunde und Überlegungen der Untersuchungen zur Reliabilität (Kapitel 2.2.3.1) und zur Validität (Kapitel 2.2.3.2), anschliessend werden einige Punkte diskutiert, die für die Interpretation der heterogenen Befundlage von Bedeutung sind (Kapitel 2.2.3.3).

2.2.3.1 Überprüfung der Reliabilität

Reliabilitätsschätzungen für Szenarien sind grundsätzlich schwierig, weil klassische testtheoretische Methoden nicht einfach übertragen werden können. *Testhalbierungs-*Schätzungen sowie interne Konsistenzprüfungen sind aufgrund der Konstruktion der Szenarien (Dynamik, Abhängigkeit des Verhaltens von früherem Verhalten) schwierig zu realisieren (einen interessanten Lösungsvorschlag hierzu macht Müller, 1993). Ähnliches gilt für die Überprüfung der *Paralleltest*-Reliabilität, die zwei strukturell in allen Aspekten gleichwertige Szenarien voraussetzt. Einige Autoren (z.B. Kluwe, Schilde, Fischer & Oellerer, 1991) sind der Überzeugung, dass es nicht möglich ist, für ein dynamisches System zwei äquivalente Formen zu entwickeln, da schon geringfügige Änderungen unterschiedliche Problemrepräsentationen bewirken. Studien zur Paralleltest-Reliabilität verwenden deshalb auch oft lediglich

ähnliche Szenarien, wobei die Ergebnisse dieser Untersuchungen dann auch unter dem Generalitätsaspekt statt unter der Reliabilitätsfrage diskutiert werden könnten (vgl. dazu auch Kersting, 1999b). *Retest*-Schätzungen schliesslich sind aufgrund von Lerneffekten ebenfalls nicht unproblematisch. Im Unterschied zu bekannten Verzerrungen bei Retest-Schätzungen anderer Verfahren (z.B. Übung, Vertrautheit mit der Testsituation) kommt bei den Szenarien der Aspekt der Problemeinsicht hinzu. So ist zu erwarten, dass beispielsweise Informationsabfragen bei einer zweiten Bearbeitung nicht mehr in der gleichen Weise nötig sind und getätigt werden wie bei der ersten Bearbeitung (aufgrund von verändertem Wissen, gemachten Erfahrungen mit der Problemsituation).

In Anbetracht der Schwierigkeit einer adäquaten Reliabilitäts-Bestimmungsmethode für ein jeweiliges Szenario, der Verschiedenheit der Ansätze und vor allem der verwendeten Indikatoren der Problemlöseleistung ist es wenig erstaunlich, dass die berichteten Korrelationskoeffizienten stark variieren: Mehrheitlich liegen die Koeffizienten im mittleren Bereich, vereinzelt werden aber auch Koeffizienten zwischen .10 und .30 sowie über .90 berichtet (vgl. Übersichten bei U. Funke, 1995; Kersting, 1999b; Kleinmann & Strauss, 1998; Strauss, 2000). Im Vergleich zu anderen eignungsdiagnostischen Verfahren zeigt sich, dass die Reliabilitäten von Szenarien deutlich unter denen von Intelligenztests liegen, jedoch mit den Reliabilitäten anderer, zum Standardrepertoire der Eignungsdiagnostik gehörenden *simulationsorientierten* Verfahren (wie Gruppendiskussion, Postkorb-Verfahren) durchaus vergleichbar sind (U. Funke, 1995, S. 189; Kersting, 1999b, S. 110; Strauss, 2000, S. 147).

Festzustellen ist, dass in neueren Studien tendenziell höhere Werte berichtet werden. Rigas, Carling und Brehmer (2002) berichten interne Konsistenzen zwischen .77 und .90 für die Szenarien „Newfire“ und „Kühlhaus“. Bezieht sich die Operationalisierung der Problemlöseleistung auf Masse des Wissenserwerbs und der Wissensanwendung, werden Werte berichtet, die nicht selten über .90 liegen (Kröner, 2001; Kröner, Plass & Leutner, 2005; Wirth, 2004; Wirth & Funke, 2005). Verschiedentlich wird auch argumentiert, dass die Reliabilitäten für Verhaltensindikatoren höher ausfallen als für Masse der Steuerungsleistung (Strohschneider, 1986; Wagener & Wittmann, 2002). Fallen Reliabilitäten für Verhaltensindikatoren gering aus, so kann

dies auf eine zu kurze Übungszeit sowie auf das anfangs instabile Verhalten einer Teilstichprobe zurückzuführen sein (Schmuck, 1992; Schmuck & Strohschneider, 1995). Reliable Messungen setzen demnach eine längere Problembearbeitung voraus, wobei die Stabilisierung des Verhaltens durch das Personenmerkmal „Verhaltensflexibilität“ moderiert wird. Personen mit hoher Kontrolleffizienz, so die beiden Studien, legen sich nicht frühzeitig auf eine bestimmte Vorgehensweise fest, sondern zeigen in den ersten Phasen der Problembearbeitung einen stärkeren Wechsel zwischen Verhaltensmustern.

Auch wenn die Reliabilität bestimmter Problemlösegütemasse unbefriedigend ist, so kann aus den vorliegenden Befunden doch geschlossen werden, dass sich bei sorgfältiger Operationalisierung der Indikatoren und Wahl der Reliabilitäts-Bestimmungsmethode zufrieden stellende Reliabilitäten finden lassen.

2.2.3.2 Überprüfung der Validität und die Intelligenzdebatte

Die wohl am häufigsten diskutierten Fragen in Studien und Arbeiten über die eignungsdiagnostische Brauchbarkeit von Problemlöseszenarien sind Fragen der Kriteriums- und der Konstruktvalidität. Zur Überprüfung der *Konstruktvalidität* werden häufig Experten-Novizen-Vergleiche herangezogen, daneben werden Vergleiche mit verwandten oder entfernten Konstrukten gemacht (konvergente, diskriminante Validität) sowie die bei verschiedenen Szenarien erzielten Leistungen miteinander verglichen (Generalitätsaspekt). Die Überprüfung der *Kriteriumsvalidität* beinhaltet mehrheitlich die für eignungsdiagnostische Instrumente fundamentale Frage der prognostischen Validität (Vorhersage für Berufserfolg) sowie – bis anhin sehr selten – die Frage der inkrementellen Validität (zusätzliche Varianzaufklärung im Berufserfolg gegenüber bestehenden Verfahren mit überlappendem Geltungsbereich).

Bei beiden Validitätsarten (Konstrukt, Kriterium) spielt das klassische psychometrische *Intelligenzkonzept* eine bedeutende Rolle. So bestand gerade in der Anfangszeit der Diagnostik mit Problemlöseszenarien die Hoffnung, mit den neuen Instrumenten

Kriterien des Berufserfolges besser vorhersagen zu können als mit klassischen Intelligenztests (die bis dahin über die höchsten Prognosevaliditäten verfügten). Deshalb wurden die vor allem in den Anfangszeiten gefundenen Nullkorrelationen zwischen Szenarien und Intelligenztests als diskriminante Validitäten interpretiert und als Beleg für eine neue Form der Intelligenz (vgl. Konzept der operativen Intelligenz von Dörner, 1986) oder zumindest für ein neues Konstrukt ausgelegt. Einen Überblick über diese ersten Studien zum Zusammenhang zwischen Intelligenz und Problemlöseleistung geben Kluwe und Mitarbeiter (Kluwe, Misiak & Haider, 1991; Kluwe, Schilde et al., 1991).

Nachfolgende Untersuchungen zum Zusammenhang zwischen Intelligenz und komplexem Problemlösen ergaben dann allerdings vermehrt auch positive und signifikante Korrelationen (vgl. zusammenfassend Süss, 1996), wobei die Frage nach den kognitiven Voraussetzungen für komplexes Problemlösen und insbesondere nach der Vorhersagekraft von Testintelligenz und bereichsspezifischem Wissen vermehrt an Bedeutung gewann. Diesbezüglich bezieht Süss (1999, S. 221) klar Stellung: „Intelligenz ist ein valider Prädiktor für komplexe Problemlöseleistungen ... [wobei] die Höhe des Zusammenhangs durch eine Vielzahl von Einflussfaktoren moderiert wird“. Anders lautende Befunde sind gemäss Süss auf die zahlreichen Methodenprobleme der entsprechenden Studien zurückzuführen. Zudem muss nach seiner Ansicht aufgrund theoretischer Überlegungen sowie empirischer Befunde die Annahme einer bereichsunspezifischen Problemlösefähigkeit, die über die bereits etablierten Konstrukte Intelligenz und Wissen hinausgeht, verworfen werden. In verschiedenen Studien kommt die Gruppe um Süss zum Schluss, dass

- die gesamte über ein Jahr stabile Problemlösevarianz (im Szenario „Schneiderwerkstatt“) durch Intelligenz und Vorwissen erklärt werden kann (Süss, Kersting & Oberauer, 1991, 1993; vgl. auch zusammenfassend Süss, 1996) und dass
- die gemeinsame Varianz der Steuerungsleistungen bei *verschiedenen* Szenarien durch Intelligenz und bereichsspezifisches Vorwissen erklärt werden kann (Kersting, 1999b, 2001; Süss, 1996; Wittmann & Süss, 1999).

Die Studie von Kersting (2001) macht zudem auch Aussagen zur (bis dahin nicht untersuchten) Frage der *inkrementellen Validität*. Der Autor untersuchte 104 überwiegend männliche Führungskräfte der Polizei. Als Prädiktoren fungierten die *Steuerungsleistungen in zwei unabhängigen Szenarien* (Berliner Version der „Schneider-

werkstatt“ sowie „DISKo“), das *systemspezifische Wissen* (für jedes Szenario einzeln erfasst) sowie *zwei Intelligenztests* (Berliner Intelligenzstrukturtest sowie ein unveröffentlichter Intelligenztest der Deutschen Gesellschaft für Personalwesen), als Kriterium die über Vorgesetztenurteile operationalisierte *berufliche Bewährung* (eigens konstruierter, umfangreicher Fragebogen zur Einschätzung der Intelligenz, Problemlösefähigkeit und Kooperationsfähigkeit). Wie in anderen Studien kann Kersting die prognostische Validität der Szenarien zwar bestätigen, nicht jedoch die inkrementelle: Ein Grossteil der kriteriumsrelevanten Varianz lässt sich durch die vorhandenen Intelligenz- und Wissensindikatoren abbilden. Daraus folgert Kersting (2001, S. 74): „Dieser Befund stellt (1) die Nützlichkeit und (2) die Konstruktvalidität computergestützter Problemlöseszenarien in Frage“. Weil die durch Intelligenztests geleistete Prognose durch die Hinzunahme des systemspezifischen Wissenstests inkrementell gesteigert werden kann, sieht Kersting die Zukunft der Szenarien mehr im Rahmen der Wissensdiagnostik. Wenn schon kein neues Konstrukt mit den Szenarien gemessen werden kann, so könnten die Szenarien immerhin für die gezielte Erfassung von Wissenserwerb und Wissensanwendung konstruiert und eingesetzt werden.

Zu einem ganz anderen Schluss kommen Wagener und Wittmann (2002), die unter Verwendung des Szenarios „FSYS“ (Wagener, 2001b) gegenüber klassischen Intelligenzmassen *zusätzliche* Varianzanteile in berufsnahen Verfahren aufklären können. Die Autoren verwendeten allerdings nicht Berufserfolgskriterien im engeren Sinn, sondern *simulationsorientierte Verfahren* (wie Interview, Gruppendiskussion, Rollenspiel, Postkorbaufgaben, Fallstudien, Präsentationen, Klausur), die im Rahmen von Assessment Centern (AC) zusammen mit dem Problemlöseszenario eingesetzt wurden (womit im Unterschied zur Studie von Kersting (2001) die konkurrente und nicht die prädiktive Validität geprüft wurde). Die Intelligenzleistung wurde (bei 59 Teilnehmern einer Personalentwicklungsmassnahme) mit der Intelligenz-Struktur-Analyse gemessen, und (bei 51 Bewerbern eines kaufmännischen Studienganges) mit dem Wilde-Intelligenztest sowie dem Berliner Intelligenzstrukturtest. Sowohl die Intelligenzmasse wie auch die FSYS-Masse korrelieren substantiell mit einzelnen AC-Übungen und einzelnen Subskalen von Postkorbaufgabe und Fallstudie sowie mit dem über alle oder ausgewählte AC-Übungen aggregierten Mass, wobei die absolu-

ten Korrelationen für FSYS vereinzelt etwas höher sind als für die Intelligenzmasse. Die Frage der inkrementellen Validität wird am Beispiel der Fallstudie und der Postkorbübung beantwortet: Auf der Ebene der FSYS-Gesamtsteuerleistung kann gegenüber den Intelligenzmassen zusätzliche Varianz aufgeklärt werden. Wird auch die Ebene der FSYS-Verhaltensskalen mitberücksichtigt, so können weitere Varianzanteile in Fallstudie und Postkorbübung aufgeklärt werden.

Diese Ergebnisse relativieren die Ergebnisse und Aussagen von Süss (1996, 1999) und Kersting (1999b, 2001): Werden nicht nur die Steuerungsleistung, sondern auch Verhaltensskalen berücksichtigt, so zeigt sich durchaus eine gegenüber Intelligenztests zusätzliche Varianzaufklärung in berufsnahen Kriterien. Damit zeigen die Szenarien gegenüber Intelligenztests einen nicht zu vernachlässigenden Zusatznutzen für die diagnostische Praxis, da doch „mehr“ gemessen wird als „nur“ (psychometrische) Intelligenz. Allerdings bleibt in der Untersuchung von Wagener und Wittmann (2002) der (mögliche) Einfluss von (Vor-) Wissen unberücksichtigt.

Genauso wie Studien zum Zusammenhang zwischen Intelligenz und Problemlöseleistung oder Studien zur prognostischen und inkrementellen Validität liefern auch die wenigen Studien zur *Generalität* der mit den Szenarien gemessenen Leistungen unterschiedliche Befunde. Nach U. Funke (1995, S. 183) liegen die berichteten Korrelationen zwischen .30 und .70, was vor allem auf die unterschiedliche Quantifizierung von Steuerungsleistungen und Verhaltensmassen in den verschiedenen Szenarien zurückzuführen sei. Eine kritischere Zusammenfassung der Befunde findet sich wiederum bei Süss (1996, S. 158f.) oder Kersting (1999b, S. 113f.). Gemäss Kersting (ebd.) liess sich die Hoffnung einiger Autoren, dass eine Gemeinsamkeit der Leistungen in verschiedenen Szenarien nachweisbar wäre, wenn vermehrt Problemlöseprozesse (Putz-Osterloh, 1983) oder System übergreifende Verhaltensweisen (Strohschneider, 1990, 1991) berücksichtigt würden, nicht überzeugend bestätigen. Allerdings, so räumt Kersting (ebd.) ein, „liegen bislang überhaupt keine konsensfähigen, psychometrisch befriedigenden Operationalisierungen von Prozessmassen des Problemlösens vor“ (S. 116). In den Zusammenfassungen von Süss (1996) oder Kersting (1999b) nicht berücksichtigt sind kleinere, wenig bekannte Studien. So zeigt beispielsweise Naef (1996, 2001a), dass sich durchaus positive Korrelationen

zwischen verschiedenen Szenarien ergeben, wenn Prozessdaten wie die Handlungs- und Vorgehensweisen statt Einzelschritte und Endzustände berücksichtigt werden.

Auch wenn Autoren wie Süss und Kersting aufgrund diverser empirischer Befunde die Annahme eines neuen Konstrukts „komplexes Problemlösen“ ablehnen und stattdessen den Einsatz der Problemlöseszenarien zum Zwecke der Intelligenz- und Wissensdiagnostik empfehlen, so sind für J. Funke (2003, 2006b) die Fragen nach der Eigenständigkeit ebenso wie die nach der Bereichsspezifität beziehungsweise Generalisierbarkeit des Konstrukts aufgrund der *heterogenen Befundlage* nach wie vor ungeklärt. Einige Punkte, die die unterschiedlichen Ergebnisse erklären können, erläutert das nächste Kapitel.

2.2.3.3 Mögliche Ursachen für die heterogene Befundlage

Einer der Hauptgründe für die verschiedenen Ergebnisse zur Reliabilität und Validität von Problemlöseszenarien liegt, wie schon an verschiedener Stelle angesprochen, in der *Vielfalt an Operationalisierungen von Problemlösegütemassen*. Einen eindrücklichen Beleg dazu liefern die Untersuchungen von Süss und Mitarbeitern: Weil die Autoren mit herkömmlichen Erfolgsmassen des Szenarios „Schneiderwerkstatt“ (Gesamtkapital bei Spielende, Anzahl der Simulationsmonate mit Gewinn) keine signifikanten Korrelationen zu Intelligenzskalen (Berliner Intelligenzstrukturtest) nachweisen konnten (Süss, Kersting und Oberauer, 1991, 1993), bildeten sie in einer Folgestudie (Süss, Oberauer & Kersting, 1993) eine neue Variable: die Summe (statt das Produkt) aus Anzahl verkauften Hemden und Gewinnspanne pro verkauftem Hemd. Und dieses Mass korrelierte signifikant mit den Intelligenzskalen.

Bei Untersuchungen zum Zusammenhang zwischen Problemlöseleistung, Intelligenz und Berufserfolgskriterien (Konstrukt-, Kriteriumsvalidität) müssen verschiedene Aspekte berücksichtigt werden. *Reliabilitätsprobleme* auf Seiten von Prädiktoren und Kriterium sowie *eingeschränkte Wertebereiche* (z.B. aufgrund nicht repräsentativer Stichproben) können dazu führen, dass die Korrelationen gering ausfallen. Dies

bestätigt eine Untersuchung von J. Funke (1983): Die Korrelation zwischen Intelligenz und Problemlöseleistung wird höher, wenn a) eine intelligenzheterogene statt -homogene Stichprobe untersucht wird, und wenn b) hoch reliable statt wenig reliable Indikatoren der Problemlöseleistung verwendet werden. Für Untersuchungen im Kontext der Intelligenzdebatte ist es deshalb wichtig, Versuchspersonen zu wählen, deren Intelligenzwerte über einen breiten Bereich streuen, statt – wie es im Rahmen psychologischer Forschung häufig vorkommt – eine Stichprobe überdurchschnittlich intelligenter Personen zu untersuchen.

Eine weitere Erklärung für gering ausfallende Korrelationen beziehungsweise die unterschiedliche Höhe der berichteten Zusammenhänge liefert die *Vernachlässigung Brunswikscher Symmetrieprinzipien*. Nach der multivariaten Reliabilitäts- und Validitätstheorie (Wittmann, 1987; Wittmann & Matt, 1986) sind hohe Korrelationen zwischen Prädiktoren und Kriterium dann zu erwarten, wenn das gleiche Generalitätsniveau vorliegt, das heisst auf beiden Seiten (Prädiktoren, Kriterium) aggregierte (globale) Werte oder auf beiden Seiten nicht aggregierte Werte vorliegen. Dies ist beispielsweise nicht der Fall, wenn die Leistung in einem Intelligenztest (als hoch aggregiertes Merkmal) mit einer spezifischen Verhaltensskala korreliert wird. Hier sind geringere Korrelationen zu erwarten, als wenn die Verhaltensskalen als (theoriegeleitetes) Aggregat mit der Intelligenz korreliert würden.

Wagener und Wittmann (2002) zeigen, dass die Zusammenhänge zwischen der Leistung im Szenario „FSYS“ (Wagener, 2001b; vgl. auch Kapitel 2.1.3), Intelligenz (Intelligenz-Struktur-Analyse) und AC-Bewertungen (Interview, Rollenspiel, Gruppendiskussion, Fallstudien, Klausur, Postkorb, Präsentationen) dann am höchsten ausfallen, wenn die AC-Bewertungen aggregiert werden, und noch höher, wenn die Aggregation theoriegeleitet erfolgt (nur Fallstudie, Postkorb, Präsentationen). Auch für die einzelnen AC-Übungen können die Autoren das Prinzip der Symmetrie bestätigen: Für die Korrelation zwischen Postkorb-Übung und Szenarienleistung beispielsweise ist die Korrelation *FSYS-Steuerleistung – Postkorb-Gesamtpunkte* höher als die Korrelationen *FSYS-Steuerleistung – Postkorb-Detailbewertungen* (einzelne Skalen), und analog sind die Korrelationen *FSYS-Verhaltensskalen – Postkorb-Detailbewertungen* höher als die Korrelationen *FSYS-Verhaltensskalen – Postkorb-Gesamtpunkte*. Substan-

tielle Korrelationen zwischen Prädiktoren (Szenarienleistung, Intelligenz) und Kriterien (AC-Übungen als Indikatoren beruflicher Leistung) können die Autoren nur bei *gleichem* Abstraktionsniveau von Prädiktoren und Kriterien nachweisen. Eine analoge Bestätigung des Symmetrieprinzips zeigt die Untersuchung zu den kognitiven Voraussetzungen komplexen Problemlösens von Wittmann und Süss (1999).

Bei Untersuchungen zum Zusammenhang zwischen Problemlöseleistungen (bzw. Intelligenz) und Berufserfolg bildet die mehrheitlich ungeprüfte und nicht selten zweifelhafte *Validität der verwendeten Berufserfolgskriterien* eine weitere Störgrösse. Bei den überwiegend verwendeten Erfolgsindikatoren wie Gehalt oder erreichte Position besteht in vielen Bereichen weder ein direkter Leistungsbezug, noch werden damit die eigentlichen beruflichen Anforderungen abgebildet (U. Funke, 1995, S. 191). Auch ergeben sich gerne Validitätsüberschätzungen, weil die Ergebnisse der Szenarienkombi bearbeitung direkt oder indirekt als Grundlage für Personalentscheidungen (wie z.B. Gehaltserhöhung oder Beförderung) dienen. Für die Erfassung von Berufserfolgskriterien fordert U. Funke (ebd.) deshalb, dass die Szenarienergebnisse den Beurteilern von Berufserfolgsdaten absolut unbekannt sind (was meistens beinhaltet, dass niemand im Unternehmen Zugang zu den Ergebnissen haben darf), und dass vermehrt arbeits- und anforderungsanalytische Ergebnisse der Operationalisierung zugrunde gelegt werden. In vielen Untersuchungen sind die Angaben über die Kriterien-erhebung zu ungenau, so dass sich nicht beurteilen lässt, ob und welche Störgrößen die Aussagekraft der Berufserfolgskriterien beeinträchtigen. Ein positives Beispiel sowohl für die Vollständigkeit der Angaben wie auch die Operationalisierung des Berufserfolges gibt die bereits erwähnte Studie von Kersting (1999b, 2001).

Die Bestimmung von Berufserfolgskriterien wird auch durch *mangelnde Kooperationsbereitschaft seitens der Unternehmen* erschwert oder gar verunmöglicht. Nicht alle Unternehmen sind bereit, fremde Beurteilungsinstrumente zu übernehmen und anzuwenden (wie in der Studie von Kersting). Bezüglich der Freigabe eigener Beurteilungsdaten sind die Unternehmen häufig sehr zurückhaltend, weil sie eine Verletzung des Datenschutzes oder sonstiger Betriebsinterna befürchten, oder entsprechende Aufzeichnungen sind überhaupt nicht vorhanden (Wagener & Wittmann, 2002, S. 83).

Faktoren, die für die unterschiedlich hohen Zusammenhänge zwischen *Problemlöseleistungen und Intelligenz* verantwortlich sein könnten, wurden bereits mehrfach diskutiert und ausführlich von Süss (1996, 1999) zusammengestellt. Auf Seite der Problemlöseleistung kritisiert der Autor vor allem die häufig *fehlende Zielspezifikation* der Problemstellung und den *ungeklärten Bezug zahlreicher Problemlösegütemasse zum jeweils spezifizierten Ziel*. Zusätzlich zu den weiter oben genannten methodischen Mängeln erwähnt Süss (1996, S. 58f.) auch

- die mangelnde Differenzierung auf Seite der Intelligenz,
- die oft fehlende Berücksichtigung von weiteren Prädiktoren wie Wissen,
- die Höhe der je nach Fragestellung zu erwartenden Effektstärken (wird neben Intelligenz auch Wissen als Prädiktor berücksichtigt, so sind nach Süss bestenfalls noch mittlere Effektstärken zu erwarten) und
- die Vernachlässigung der Beziehung zwischen Effektstärke und optimalem Stichprobenumfang.

Die verschiedentlich gemachte Aussage, viele Untersuchungen hätten keinen Zusammenhang zwischen Problemlöseleistung und Intelligenz nachweisen können, ergänzt Süss (1996, S. 59) kritisch mit dem Hinweis, „dass diese Untersuchungen dazu auch gar nicht in der Lage waren“.

Neben methodenkritischen Überlegungen liefern auch theoretische Überlegungen zur Beziehung zwischen Intelligenz und komplexem Problemlösen mögliche Erklärungen für die heterogene Befundlage. Nach Leutner (2002) lassen sich die unterschiedlichen und nicht selten widersprüchlichen Befunde anhand der so genannten *Elshout-Raaheim-Hypothese* erklären (Elshout, 1987; Raaheim, 1988). Diese postuliert eine umgekehrt u-förmige Beziehung zwischen dem Ausprägungsgrad des Korrelationskoeffizienten (zwischen Intelligenz und Problemlösen) als abhängige Variable und dem Ausmass an verfügbarem bereichsspezifischem Vorwissen als unabhängige Variable. Das heisst: Bei mittlerem Vorwissen über das Problem ist der Zusammenhang zwischen Intelligenz und Problemlösen maximal, bei geringem Vorwissen (und damit in der Anfangsphase der Problembearbeitung) und hohem Vorwissen (nach längerer Problembearbeitung) ist der Zusammenhang jedoch gering. In zwei Experimenten mit dem Szenario „Hunger in Nordafrika“ kann Leutner (ebd.) diese Hypothese mehrheitlich bestätigen. Die Resultate des ersten Experiments (between-

subject-Design mit je 100 Studierenden und Schülern und Manipulation des Vorwissens) entsprechen fast vollständig dem postulierten Korrelationsmuster, die Resultate des zweiten Experiments (within-subject-Design mit 28 jungen Erwachsenen und je drei Messzeitpunkten an drei aufeinanderfolgenden Tagen sowie einer Messung nach zwei Wochen) nur bedingt: Das postulierte umgekehrt u-förmige Korrelationsmuster zeigt sich nicht über alle zehn Messzeitpunkte, sondern nur innerhalb der drei Messungen an einem Tag. Die Resultate bestätigen aber dennoch die Vermutung, dass der Vorhersagewert der Intelligenz für komplexe Problemlöseleistungen abhängig ist vom Grad des Wissens der Person, welches sich über die Zeit hinweg verändern kann. Intelligenz spielt für die Problemlösung nach diesen Befunden nur dann eine Rolle, wenn die Person durch vorausgegangene Systemexploration bereits Gelegenheit zur Wissensaneignung gehabt hat und dieses Wissen nun aneignen möchte.

Ein aufwändigeres Modell zum Zusammenhang zwischen (Vor-) Wissen, Intelligenz und Problemlöseleistung wurde von Kröner (2001) vorgeschlagen und in der Folge studie von Kröner und Mitarbeitern (2005) erneut empirisch bestätigt. Nach diesem Modell gibt es eine unabhängige Wirkung der Testintelligenz auf die mit dem Szenario „Multiflux“ gemessenen Variablen Explorationsstrategie, Systemwissen (Wissenserwerb) und Wissensanwendung (vgl. Kapitel 2.1.2). Demnach verfügen intelligentere Personen im Vergleich zu weniger intelligenten Personen über eine bessere Explorationsstrategie, eignen sich mehr Systemwissen an und sind erfolgreicher in der Wissensanwendung (Systemsteuerung). Eine erfolgreiche Explorationsstrategie ist dabei Voraussetzung für die Aneignung von Systemwissen und dieses wiederum ist Voraussetzung für eine erfolgreiche Wissensanwendung. Aufgrund der substantiellen Korrelationen zwischen Intelligenz (Berliner Intelligenz-Strukturtest) und den drei Variablen beziehungsweise einem aus diesem aggregierten Mass sehen die Autoren in „Multiflux“ damit sogar ein zur Intelligenzdiagnostik *alternatives* Instrument.

Im Hinblick auf die vielschichtige Diskussion rund um den Zusammenhang zwischen Intelligenz und komplexem Problemlösen sollten schliesslich zwei Punkte nicht unbeachtet bleiben:

1. Intelligenz stellt selber ein sehr breites Konstrukt dar, wovon die psychometrische Intelligenz nur einen (wiederum breiten) Teilaspekt bildet. Schweizer (2006b) gibt eine kurze Übersicht über verschiedene *psychometrische* Intelligenztests, und Weber und Westmeyer (2001) diskutieren (kritisch) *alternative* Intelligenzkonzepte.
2. Intelligenz gilt nach wie vor als der beste Prädiktor für beruflichen Erfolg (Schweizer, 2006a). Andere eignungsdiagnostische Verfahren sollten – der zur Zeit umfangreichsten Meta-Analyse über die Vorhersage von Berufsleistung zufolge – nur in Ergänzung zu klassischen Intelligenztests angewendet werden, mit dem Ziel inkrementelle Validität zu erzeugen (Schmidt & Hunter, 1998a, 1998b).

In diesem Sinne muss man J. Funke (1998, S. 94) zustimmen, der schreibt: „PC-based simulations will not replace other diagnostic instruments. Rather, they enrich the existing instruments with new aspects“. Computergestützte Problemlöseszenarien sollen die klassische Intelligenzdiagnostik *ergänzen*, nicht ersetzen. Die Hoffnung, Szenarien könnten im Vergleich zu klassischen Intelligenztests als die „besseren“ Instrumente abschneiden, sollte begraben werden. Im Hinblick auf die Erfüllung und Prüfung eignungsdiagnostischer Qualitätsanforderungen sind Intelligenztests den Szenarien eindeutig überlegen. Angesichts der Vorteile und Nutzen der Szenarien für die eignungsdiagnostische Praxis ist es jedoch erstaunlich, dass einige Autoren dafür plädieren, Szenarien gezielt und ausschliesslich für die Intelligenzdiagnostik beziehungsweise die Diagnose von Wissenserwerb und Wissensanwendung zu konstruieren (Kersting, 2001; Kröner, 2001; Kröner et al., 2005). Damit stünde der diagnostischen Praxis nicht mehr als ein weiteres Verfahren zur (Status-) Diagnostik intellektueller Fähigkeiten zur Verfügung. Die Möglichkeit einer *Erweiterung* herkömmlicher Intelligenzdiagnostik bleibt bei dieser Perspektive unberücksichtigt. Das Potential der Szenarien zur Prozessdiagnostik, die Möglichkeit detaillierte Angaben über Bearbeitungsprozesse und verschiedenste Aspekte des Verhaltens zu machen, unterscheidet die Szenarien deutlich von Intelligenztests. Diese grosse Menge zusätzlicher Information, die Szenarien – bei sorgfältiger Konstruktion und Operationalisierung – zur Verfügung stellen *können*, sollte nicht einfach aufgrund einiger enttäuschender Befunde oder kritischer Stimmen ignoriert werden.

2.3 Folgerungen und die Beschreibung eines neuen Auswertungsmodells

Gerade der zuletzt genannte Punkt, nämlich die Möglichkeit einer differenzierten Verhaltens- und Prozessdiagnostik, wird von Befürwortern einer Diagnostik mit Problemlöseszenarien als der entscheidende Vorteil gegenüber herkömmlichen eignungsdiagnostischen Verfahren genannt. Dieser Argumentationslinie folgte bis hierher auch diese Arbeit. In Anlehnung an Kersting (1999b) müssen jedoch einschränkend drei Punkte betrachtet und weiter vertieft werden:

1. Die bei vielen Szenarien fehlende oder mangelnde Standardisierung der Auswertung stellt die diagnostische Brauchbarkeit der Szenarien in Frage.
2. Hinsichtlich einer auf Problemlöseszenarien aufbauenden Prozessdiagnostik herrscht keinerlei Konsens. Stattdessen wird eine Vielzahl unterschiedlichster Verhaltens- und Prozessmasse vorgeschlagen.
3. Die Vielzahl unterschiedlichster Problemlösegütemasse erschwert eine vergleichende Interpretation der Befunde.

Ad 1: *Fehlende oder mangelnde Standardisierung der Auswertung.* Oft ist unklar, wie die verschiedenen, oft widersprüchlichen Indikatoren der Problemlöseleistung interpretiert und zu einem Gesamturteil integriert werden müssen. Offen bleibt auch, welchen Alltagsbezug einzelne Verhaltensskalen haben (wie relevant ist beispielsweise das Merkmal „Anzahl gestellter Warum-Fragen“?). Kersting (1999b, S. 69) bringt es auf den Punkt: „Häufig können lediglich die Programmierer des Problemlöseszenarios sowie wenige Experten, nicht aber die diagnostischen Anwender oder die Problemlöser selbst, nachvollziehen wie die einzelnen Indizes zustande kommen“. Damit haftet dem Urteil des Diagnostikers eine grosse Willkür an. In Bezug auf die Interraterreliabilität dürften kaum zufrieden stellende Ergebnisse berichtet werden. Aber auch bei ein und demselben Diagnostiker kann nicht davon ausgegangen werden, dass er bei unterschiedlichen Anwendungen die gleichen impliziten Regeln zur Datenintegration anwendet. Um wenigstens annähernd konsistente Urteile zu erzielen, muss sich ein Diagnostiker Expertenwissen aneignen, was Zeit und letztlich auch Kosten intensiv ist. Dabei würde es sich ja geradezu anbieten, den ohnehin vorhandenen Computer auch für eine Automatisierung der Datenintegration und – soweit möglich – der Dateninterpretation zu nutzen. Ein

einfach handhabbares, leicht verständliches Auswertungsschema, das (ähnlich der Ergebnisausgabe bei Intelligenztests) eine rasche und eindeutige Einschätzung der Problemlösefähigkeit ermöglicht, wäre wünschenswert.

Ad 2: Fehlender Konsens über eine auf Problemlöseszenarien aufbauende Prozessdiagnostik. Über die Unterschiedlichkeit der vorgeschlagenen Operationalisierungen von Verhaltens- und Prozessmassen wurde bereits in den Kapiteln 2.1.2 und 2.1.3 berichtet. Bei kritischer Betrachtung zeigt sich zudem, dass nur wenige Ansätze wirklich Aussagen über den *Prozess* der Bearbeitung erlauben. Wie bei Massen der Steuerungsleistung steht bei vielen Anwendungen das *Endprodukt* im Vordergrund: das Wissen, das sich die Versuchsperson während der Bearbeitung angeeignet hat, die Anzahl gestellter Fragen, getätigter Massnahmen, die Zeit für Entscheidungen, Art und Gegenstand von Entscheidungen usw. Einzelne Verhaltens- (oder Wissens-) Aspekte werden ausgewählt und ausgezählt oder beschrieben und beurteilt und als *globale Masse* abgebildet. Nur sehr vereinzelt werden *Verlaufs- und Prozessmasse* im eigentlichen Sinne gebildet, indem einzelne Verhaltensaspekte in ihrem zeitlichen Verlauf (meist über einzelne Takte) dargestellt (z.B. Strohschneider, 1994) oder der Prozess des Wissenserwerbs und der Wissensanwendung abgebildet werden (z.B. Wirth, 2004; Wirth & Funke, 2005).

Verhaltens- und Prozessorientierung kann aber mehr bedeuten als nur *ausgewählte* Verhaltens- (oder Wissens-) Aspekte zu berücksichtigen: Da der Einsatz des Computers die Registrierung aller von der Versuchsperson getätigten Eingriffe und Arbeitsschritte (in der Regel Mausklicks) ermöglicht, sollten auch alle diese einzelnen Aktionen für die Auswertung und Beurteilung der Problemlöseleistung herangezogen werden. Im Folgenden werden diese automatisch registrierten Eingriffe der Versuchsperson als *Mikroverhalten* bezeichnet, weil sie die kleinsten, automatisch registrierbaren Verhaltenseinheiten darstellen. Prozessorientierung bezieht sich auf zwei Ebenen: auf die Bewertung jedes dieser Mikroverhalten *in seinem aktuellen Handlungskontext* (im Hinblick auf für die Problemlösung als wichtig erachteten Aspekte) und auf die Bewertung des *Problemlöseprozesses als Ganzes*.

Ad 3: *Vielzahl an Operationalisierungen erschwert vergleichende Interpretation.* Schon im vorangegangenen Kapitel wurde erwähnt, dass die Vielfalt der vorgeschlagenen Problemlösegütemasse einer der Hauptgründe für die verschiedenen Ergebnisse zur Reliabilität und Validität der Problemlöseszenarien ist. Für jedes Szenario werden andere Indikatoren der Problemlöseleistung herangezogen, wobei es sich mit wenigen Ausnahmen um *szenariospezifische* Masse handelt. Szenariospezifisch meint dabei, dass die Masse abhängig sind vom Inhalt des jeweiligen Szenarios und damit nur für dieses Szenario gelten und nicht auf andere Szenarien übertragen werden können, oder wenn doch, dann nur auf Szenarien mit ähnlichem Inhalt. Beispielsweise lassen sich die Verhaltensskalen „Kontinuierliche Kontrolle des Systems“ und „Gleichmässige Kontrolle“ des Szenarios „FSYS“ (Wagener, 2001b) nur auf Szenarien übertragen, die ebenfalls die eigenständige Kontrolle des Systemzustandes durch den Probanden erfordern. Die „Art und Gegenstand der Fragen“ (Dörner, 1986) lässt sich nur operationalisieren, wenn das Stellen von Fragen durch die Versuchsperson überhaupt vorgesehen ist (etc.). *Szenario-unspezifische* Masse demzufolge sind vom Inhalt unabhängige Masse und können für verschiedene Szenarien in analoger Weise berechnet werden. Das bedeutet, dass ihre Einschätzung hinsichtlich allgemeiner Konstrukte (und nicht spezifischer Aspekte) erfolgen muss, die selbstverständlich für das Lösen komplexer Probleme grundlegend sein müssen. Ein Beispiel ist die Erfassung von Wissenserwerb und Wissensanwendung, wie sie in Kapitel 2.1.2 vorgestellt wurde (z.B. Kröner, 2001; Wirth, 2004; Wirth & Funke, 2005). In analoger Weise lassen sich auch bei anderen (vergleichbaren) dynamischen Systemen Wissenserwerb und -anwendung implementieren.

Durch szenario-unspezifische Masse werden die Leistungen in verschiedenen Szenarien direkter vergleichbar, weil „Gleiches mit Gleichem“ verglichen wird. Es ist zu erwarten, dass damit auch einheitlichere Befunde hinsichtlich Reliabilität und Validität der Instrumente gewonnen werden können.

Im Folgenden wird ein neues Auswertungsmodell vorgestellt, das die eben geschilderten drei Aspekte berücksichtigt. Endprodukt der neuen Auswertung ist ein automatisierbares Auswertungs- und Beurteilungsschema, das sowohl die Dateninterpretation wie auch die Datenintegration standardisiert. Das Schema ist einfach

anwendbar und verständlich, so dass auch ohne Expertenwissen eine rasche und treffsichere Beurteilung der Probandenleistung möglich ist. Grundlage der Leistungsbeurteilung ist die *prozessorientierte* Einschätzung von *szenario-unspezifischen* Konstrukten auf der Basis von *Mikroverhalten*.

Als erstes muss erklärt werden, welche szenario-unspezifischen Konstrukte definiert werden und warum. Charakteristisch für komplexe Probleme ist, wie in der Einleitung erläutert, dass die für die Problemlösung relevanten Informationen zu Beginn grösstenteils nicht verfügbar sind, sondern erst im Verlaufe der Problembearbeitung erschlossen werden müssen. Informationssuche, Informationsaufnahme und Informationsintegration sind zentrale Bestandteile einer erfolgreichen Problembearbeitung. Oder aus der Perspektive der Wissensdiagnostik (vgl. Kapitel 2.1.2): Zum einen muss Wissen über das System und die Möglichkeiten seiner Steuerung interaktiv erworben werden, zum anderen muss dieses Wissen angewendet werden, um das System erfolgreich zu steuern. Es würde sich also anbieten, auf die vorgeschlagenen Methoden zur Erfassung von Wissenserwerb und Wissensanwendung zurückzugreifen, da damit szenario-unspezifische Konstrukte eingeschätzt werden und erst noch verschiedene Studien die Bedeutung von Wissenserwerb und Wissensanwendung für erfolgreiches Problemlösen belegen (vgl. zusammenfassend J. Funke, 2003, S. 158-159 und S. 163-170). Ihre Erfassung beschränkt sich jedoch auf Szenarien, die auf der Basis formaler Modelle konstruiert wurden, während bei inhaltlich orientierten Szenarien nur das Vorwissen oder der Wissenserwerb zuverlässig operationalisiert werden kann (vgl. Kapitel 2.1.2).

Deshalb werden in dieser Arbeit zwei alternative szenario-unspezifische Konstrukte vorgeschlagen: Informationseröffnendes Handeln und Adäquates Handeln. *Informationseröffnendes Handeln* (abgekürzt: Info) meint, inwiefern die Versuchsperson mit jedem Handlungsschritt neue, bezüglich der aktuellen Problemstellung relevante Information aufnimmt. *Adäquates Handeln* (abgekürzt: Ad) meint, inwiefern der jeweilige Handlungsschritt im aktuellen Handlungskontext und in Bezug auf die aktuelle Problemstellung sinnvoll ist. Adäquates Handeln setzt (zumindest teilweise) die Aufnahme und das Umsetzen von relevanten Informationen voraus. Um sich einer sich dauernd verändernden Problemsituation anzupassen und sich so zu

verhalten, dass man das vorgegebene Ziel erreicht, muss man über die relevanten Informationen verfügen und diese (implizit oder explizit) verstehen. Die beiden Konstrukte, die im Methodenteil näher beschrieben werden, sind mit den Konzepten Wissenserwerb und Wissensanwendung vergleichbar. Sie beziehen sich jedoch nur auf die *potentielle* Informationsaufnahme und damit auf potentielles Wissen, da ihre Einschätzung auf reinen (vom Programm registrierten) Verhaltensdaten beruht. Vorausgesetzt, das Programm speichert in einer Protokolldatei alle Eingriffe der Versuchsperson zeitsynchron ab, so sollte eine Einschätzung des Informationseröffnenden und Adäquaten Handelns grundsätzlich für jedes Szenario – unabhängig von Inhalt und Struktur – möglich sein.

Abbildung 1 zeigt, aus welchen Bausteinen das neue Auswertungsmodell besteht. Die Beschreibung des Modells kann an dieser Stelle nur sehr abstrakt erfolgen. Dargestellt werden soll der Grundgedanke des Modells, das Gerüst sozusagen. Eine exakte und inhaltsbezogene Beschreibung erfolgt im Methodenteil der Arbeit.

Kernstück der neuen Auswertung bildet die *prozessorientierte Mikroverhaltensbeschreibung*. Jedes vom Programm registrierte Mikroverhalten (= Arbeitsschritte/Aktionen des Probanden, in der Regel registrierte Mausklicks) wird – nach bestimmten, entsprechend der Zielsetzung des Szenarios zu definierenden Kriterien – im Hinblick auf das Informationseröffnende und Adäquate Handeln auf einer Ratingskala beurteilt. Als Resultat der Mikroverhaltensbeschreibung liegen dann für jede Problembearbeitung zwei Zahlenfolgen vor, das heisst zeitliche Abfolgen von *Mikroverhaltensdaten*. Da der Prozess der Problembearbeitung interessiert, werden diese Mikroverhaltensdaten über die Zeit aufsummiert und als *Zeitreihen* graphisch dargestellt. Für jeden Probanden liegen damit zwei Zeitreihen vor, die aufzeigen, wie sich der Proband während der Problembearbeitung in Bezug auf das Informationseröffnende und Adäquate Handeln verhält (vgl. Graphik in Abbildung 1). Zwei steil ansteigende Kurven beispielsweise bedeuten, dass sich der Proband mit jedem Schritt neue Informationen verschafft und sich in Bezug auf die Problemlösung optimal und effizient verhält.

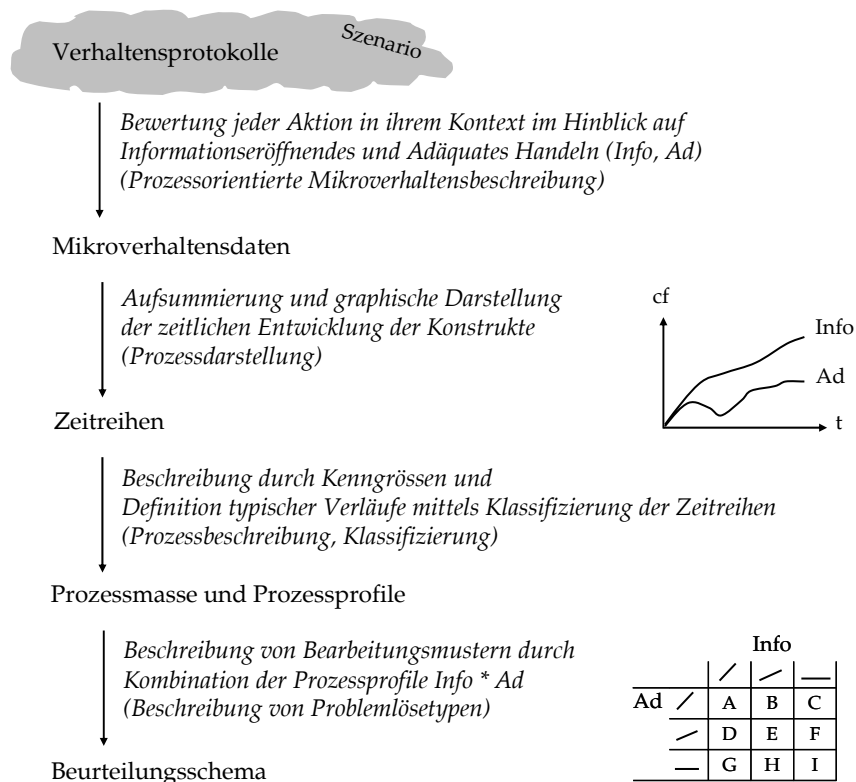


Abbildung 1: Übersicht über das neue Auswertungsmodell. Die einzelnen Bausteine werden im Text näher erläutert.

Diese Zeitreihen sollen nun so beschrieben werden, dass ein quantifizierbares Beurteilungsschema der Probandenleistung resultiert. Zum einen sollen für jedes Konstrukt Kenngrößen definiert werden, die den Kurvenverlauf *formal beschreiben*. Diese Kenngrößen bilden ein *Prozessmass* und grenzen sich damit von anderen verhaltensbasierten globalen Massen ab. Zum anderen können für jedes Konstrukt (unter Einbezug formaler Klassifizierungsmethoden) typische, charakteristische Verläufe definiert werden, die die *inhaltliche Bedeutung der Verläufe* und damit die *Bewertung der Problembearbeitung* qualitativ beschreiben. Diese typischen Verläufe für das Informationseröffnende und Adäquate Handeln werden als *Prozessprofile* bezeichnet. Ziel ist, für jedes Konstrukt wenige, klar voneinander abgrenzbare (disjunkte), hierarchisch geordnete Prozessprofile zu definieren.

Die Kombinationen der Prozessprofile der beiden Konstrukte, dargestellt in einer Kreuztabelle, ergeben das *Beurteilungsschema*. Jedes Feld beschreibt ein bestimmtes Bearbeitungsmuster, definiert durch die typischen Verläufe (Prozessprofile) für das

Informationseröffnende und Adäquate Handeln. Diese typischen Problembearbeitungen werden als *Problemlösetypen* bezeichnet. Aufgrund der ordinalen Information der Prozessprofile kann definiert werden, welche Problemlösetypen „geeignet“ sind und welche weniger oder gar nicht. Für das in Abbildung 1 eingezeichnete Beispiel eines Beurteilungsschemas wäre denkbar, die Typen A, B, D, E als gute bis sehr gute Problemlöser zu bezeichnen.

Das Resultat der neuen Auswertung ist damit denkbar einfach: Die Leistung einer Person wird eingeschätzt durch die Zuordnung zu einem bestimmten Feld auf dem Beurteilungsschema, durch die Zugehörigkeit zu einem bestimmten Problemlösetypen, der eine typische Problembearbeitung beschreibt. Diese Zuordnung ermöglicht eine rasche (und standardisierte!) Aussage über die Güte der Problemlöseleistung. Durch die Berücksichtigung der *individuellen Zeitreihen* sind anschliessend differenzierte Interpretationen des Bearbeitungsprozesses möglich.

2.4 Fragestellungen und Hypothesen

Ziel der vorliegenden Arbeit ist, das vorgestellte theoretische Auswertungsmodell auf ein Szenario anzuwenden, das im Rahmen der Personalauswahl eingesetzt wird.

Als erstes muss überprüft werden, ob der neue Ansatz die individuellen Unterschiede im Problemlöseverhalten beschreiben kann und ob sich für jedes Konstrukt wenige, klar voneinander abgrenzbare Prozessprofile definieren lassen. Da die Definition der Prozessprofile in einem ersten Schritt auf inhaltlichen Überlegungen basiert, wird zudem geprüft, ob sich diese qualitative Klassifizierung der Zeitreihen durch formale Klassifizierungsmethoden bestätigen lässt. Positive Befunde zu diesen Fragestellungen bilden die Voraussetzung dafür, dass die folgenden Fragestellungen zur Überprüfung des neuen Ansatzes überhaupt untersucht werden können.

Um erste Aussagen über die Validität des neuen Ansatzes zu gewinnen, werden drei unterschiedliche Aspekte berücksichtigt. Als erstes wird der neue Ansatz mit der

vorhandenen, praxisbewährten, *szenariospezifischen Auswertung* verglichen, um sicher zu stellen, dass die neue Auswertung nicht etwas ganz anderes misst als die bestehende Auswertung. Es wird erwartet, dass die Einschätzungen der Problemlöseleistung durch die beiden Ansätze in die gleiche Richtung gehen.

Weiter wird die neue Auswertung einem *Expertenurteil* gegenübergestellt. Dieser Experte soll die Aufgabe eines Diagnostikers übernehmen, der die verschiedenen und zum Teil widersprüchlichen Daten, die nach der Bearbeitung des Szenarios vorliegen, zu einem Gesamturteil integriert. Dieses Gesamturteil soll zudem szenario-unspezifisch sein, das heisst die szenariospezifischen Ergebnisse müssen „übersetzt“ werden in eine allgemeine Einschätzung der Problemlösefähigkeit der Probanden. Der Experte erhält die Aufgabe, die Probandenleistung in Bezug auf die durch das Auswertungsmodell definierten Konstrukte (Info, Ad) beziehungsweise deren Prozessprofile einzuschätzen. Die Prüfung der Übereinstimmung der beiden Urteile geht der Frage nach, wie gut das neue Auswertungsmodell die aufwändige Datenintegration und -interpretation durch den Diagnostiker (den Experten) ersetzen kann.

Als drittes wird überprüft, ob sich auch an einem *unabhängigen Datensatz* eine Übereinstimmung zwischen neuer Auswertung und Expertenurteil nachweisen lässt. Mit unabhängigem Datensatz ist eine Stichprobe gemeint, die nicht zur Entwicklung des Beurteilungsschemas (Mikroverhaltensbeschreibung, Definition der Prozessprofile durch Klassifizierung der Zeitreihen) beigetragen hat. Damit soll ausgeschlossen werden, dass die Aussagen nur für die untersuchte Stichprobe Gültigkeit haben.

Für eine bessere Übersicht werden die sechs Fragestellungen (und wo definierbar auch die statistischen Hypothesen) im Folgenden zusammenfassend dargestellt und durchnummeriert. Diese Nummerierung beziehungsweise Reihenfolge wird bei der Beschreibung der Auswertungsverfahren (Kapitel 3.5) sowie bei der Darstellung und Diskussion der Ergebnisse (Kapitel 4 und 5) wieder aufgenommen.

Fragestellungen 1-3: Prozessdarstellung und -beschreibung

1. Kann das neue Auswertungsmodell (anhand der zeitlichen Verläufe der beiden Konstrukte) die individuellen Unterschiede im Problemlöseverhalten beschreiben?
2. Lassen sich für die zeitlichen Verläufe der beiden Konstrukte wenige, klar voneinander abgrenzbare Prozessprofile (disjunkte Klassen) definieren?
3. Lässt sich die inhaltlich begründete, qualitative Klassifizierung der Zeitreihen (d.h. die Definition der Prozessprofile) durch formale Klassifizierungsmethoden bestätigen?

Fragestellungen 4-6: Überprüfung des Auswertungsmodells

4. Wird die Problemlöseleistung durch das neue, prozessorientierte Auswertungsmodell ähnlich eingeschätzt wie durch den szenariospezifischen Auswertungsansatz?

Hypothese 4.1: Die *Prozessmasse* für das Informationseröffnende und Adäquate Handeln stehen mit den szenariospezifischen Leistungsmassen in einem positiven Zusammenhang.

Hypothese 4.2: Die *Prozessprofile* für das Informationseröffnende und Adäquate Handeln unterscheiden sich im Hinblick auf die szenariospezifischen Leistungsmasse, wobei das beste Profil die höheren Werte aufweist als das zweitbeste Profil und dieses wiederum die höheren Werte als das drittbeste Profil etc. (monoton fallender Trend).

5. Kann der neue Ansatz die Datenintegration und -interpretation durch den Diagnostiker ersetzen?

Hypothese 5.1: Die Einschätzungen der Problemlöseleistung durch das neue Auswertungsmodell (d.h. durch die Prozessprofile) stimmen mit dem Urteil des Experten überein.

6. Lässt sich das neue Auswertungsmodell auch an einem von der Entwicklung unabhängigen Datensatz anwenden?

Hypothese 6.1: Das neue Auswertungsmodell (Kriterien der Mikroverhaltensbeschreibung, Zuordnung zu Prozessprofilen) lässt sich auch an einem unabhängigen Datensatz ohne Einschränkung anwenden.

Hypothese 6.2: Auch für einen unabhängigen Datensatz stimmen die Einschätzungen der Problemlöseleistung durch das neue Auswertungsmodell (d.h. durch die Prozessprofile) mit dem Urteil des Experten überein.

3 Methodisches Vorgehen

Dieser Teil der Arbeit legt dar, wie die eben erläuterten Fragestellungen beantwortet werden. Kapitel 3.1 stellt das Problemlöseszenario vor, das für die Anwendung und Überprüfung des neuen Auswertungsmodells verwendet wird. Kapitel 3.2 beschreibt die schrittweise Entwicklung des Auswertungsmodells und Kapitel 3.3 erläutert das Zustandekommen des Expertenurteils. In den letzten beiden Kapiteln folgen Angaben zum Vorgehen und zur Stichprobe (Kapitel 3.4) sowie zu den Auswertungsverfahren (Kapitel 3.5).

3.1 Anwendungsbeispiel: Das Szenario ATLANTIS

Das neue Auswertungsmodell soll auf ein Problemlöseszenario angewendet werden, das in der Personalselektion eingesetzt wird. Deshalb muss das Szenario bestimmte Anforderungen erfüllen. In Kapitel 2.2.3 wurde bereits erläutert, welchen *Qualitätsanforderungen* die Szenarien genügen müssen, wenn sie in der Eignungsdiagnostik eingesetzt werden sollen. Es wurde auch darauf hingewiesen, dass die meisten Szenarien die von Hasselmann (1995, S. 239) gestellten Anforderungen an die *Szenarienkonstruktion* (Punkte 1-4) nicht erfüllen, und die Anforderungen im Hinblick auf die *eignungsdiagnostische Anwendung* der Szenarien (Punkte 5-9) nur vereinzelt geprüft werden. Kröner und Mitarbeiter (2005) fordern für die *Szenarienkonstruktion* zudem die Berücksichtigung von drei weiteren Aspekten:

1. Den Ausschluss von szenariospezifischem Vorwissen (vgl. dazu auch die Überlegungen in Kapitel 2.1.2).
2. Die Möglichkeit zur Exploration des Systems durch Vorgabe einer Übungsphase.
3. Das Herstellen vergleichbarer Bedingungen für alle Probanden durch Vorgabe eines klaren Ziels und einer gut definierten optimalen Lösung.

Eine weitere Anforderung stellt die *Mikroverhaltensbeschreibung* des neuen Auswertungsmodells: Das Verhalten muss zeitsynchron registriert werden und für jedes Mikroverhalten muss auch der Handlungskontext gespeichert werden.

Mit der Wahl des Szenarios ATLANTIS (Hirsig & de With, 1996, 1998, 2001) können die erwähnten Anforderungen an *Szenarienkonstruktion* und *Datenspeicherung* erfüllt werden. Wie für die meisten Szenarien liegen jedoch keine Angaben über die test-theoretischen Gütekriterien vor. Das Szenario wird aber seit mehreren Jahren in einem grossen Schweizerischen Dienstleistungsunternehmen³ für die Personalauswahl und -entwicklung eingesetzt, was zumindest ansatzweise für die *eignungsdiagnostische Brauchbarkeit* des Instruments spricht.

Das Szenario ATLANTIS wird in den folgenden Abschnitten vorgestellt. Nach einer Übersicht über Aufbau und Inhalt des Instruments (Kapitel 3.1.1 und 3.1.2) folgen Angaben zum bestehenden, szenariospezifischen Auswertungsansatz (Kapitel 3.1.3).

3.1.1 Der simulierte Realitätsausschnitt

Das Szenario ATLANTIS versetzt den Probanden auf ein Schiff einer Forschungsflotte, die im Atlantik Untersuchungen im Bereich der Biologie und der Geologie ausführt. Die konkrete Aufgabe des Probanden besteht darin, in einem festgelegten Gebiet mittels eines ferngesteuerten Mini-U-Bootes die genaue Lage von erzhaltigem Gestein für künftige Probebohrungen festzustellen und zu kartographieren. Diese klar umschriebene Aufgabenstellung dient dem Kennenlernen der zur Verfügung stehenden Instrumente und bildet den Einstieg in ein spannendes Szenario: Kurze Zeit nachdem sich der Proband mit der Erzsuche vertraut gemacht hat, wird ein Passagierflugzeug als vermisst gemeldet und alle Einheiten der Forschungsflotte werden aufgefordert, sich umgehend an der Suche nach dem vermutlich abgestürzten Flugzeug beziehungsweise der Bergung möglicher Überlebender zu beteiligen.

Der Proband sitzt am Bildschirm des Steuerungscomputers für das unbemannte Mini-U-Boot Nautilus. Dieser Steuerungscomputer mit dem Namen Nautilus EA.4 umfasst vier Funktionseinheiten (vgl. Abbildung 2):

³ Aus Datenschutzgründen möchte das Unternehmen anonym bleiben.

1. Eine Bedienungseinheit für die *Fernsteuerung* von Nautilus. Für die *Fortbewegung* von Nautilus stehen vier Modi zur Verfügung: horizontales Fahren in acht Richtungen, Sinken, Steigen, automatisch Auftauchen bis zur Meeresoberfläche und automatisch Absinken bis drei Meter über Meeresgrund (wobei dieser Abstand bei der horizontalen Weiterfahrt beibehalten wird, so dass Kollisionen mit dem Grund verhindert werden). Der Proband kann diese Modi nur einzeln aufrufen, das heisst er kann Nautilus nur schrittweise in horizontaler oder vertikaler Richtung bewegen. Die Fernsteuerung von Nautilus ermöglicht zudem den Einsatz eines Metall-Detektors, das so genannte *Magneto-Sonargerät*, das auf Knopfdruck eisenhaltiges Material im Umkreis von 20 Metern wie auf einem Radarschirm anzeigt, sowie den Einsatz einer *Videokamera*, die auf Wunsch einzelne Bilder vom aktuellen Sichtwinkel (Sichtbereich ca. vier Meter) übermittelt.
2. Ein *Kommunikationssystem*, über das der Proband mit anderen Mitarbeitern oder der Leitung des Forschungsteams schriftlich kommunizieren kann. Das Kommunikationssystem beinhaltet – ähnlich einem E-Mailprogramm – zwei Funktionen: Über die *Funktion Message lesen* kann die jeweils zuletzt eingegangene Nachricht aufgerufen werden. Über die *Funktion Protokoll/Message* kann der Proband selber zu jedem Zeitpunkt Nachrichten an andere Sucheinheiten senden. Zudem wird er aufgefordert, gelegentlich den aktuellen Arbeitsstand und seine Überlegungen dazu in Form von Protokollen – für sich und für andere Mitarbeiter – zu dokumentieren.
3. Eine *Datenbank* mit dem Namen *Nauti-Info*, die Informationen über verschiedene Aspekte der Schifffahrt enthält (wie z.B. Informationen über Meeresbiologie, Geologie und Strömungen oder nautische Instrumente). Die verschiedenen Informationen sind für die Problemlösung unterschiedlich relevant.
4. Eine *Bedienungshilfe* mit Erklärungen zu allen Funktionen des Steuerungscomputers. Wird die Hilfe eingeschaltet, verbleibt das Drücken aller Knöpfe ohne Wirkung, stattdessen wird jeweils ein erklärender Text eingeblendet.

Der Steuerungscomputer ist so konzipiert, dass Handlungen nicht parallel laufen können (kein „multi-tasking“), das heisst der Proband kann nicht eine Nachricht verfassen und gleichzeitig in der Datenbank etwas nachschlagen oder noch schnell ein Videobild anfordern. Die Nachricht muss zuerst verschickt (oder verworfen) werden, dann kann die Datenbank geöffnet werden. Eine detaillierte Übersicht über alle verfügbaren Funktionen beziehungsweise den gesamten Menü-Aufbau des Steuerungscomputers findet sich in Anhang B.

Neben den Eingaben via Bildschirm erhält der Proband zudem die Möglichkeit, sein Vorgehen und seine Überlegungen auf einer *Seekarte* zu notieren. Die Karte zeigt analog einer Landkarte das aktuelle Forschungsgebiet sowie den Standort des Forschungsschiffes, von dem aus operiert wird (vgl. Anhang B).

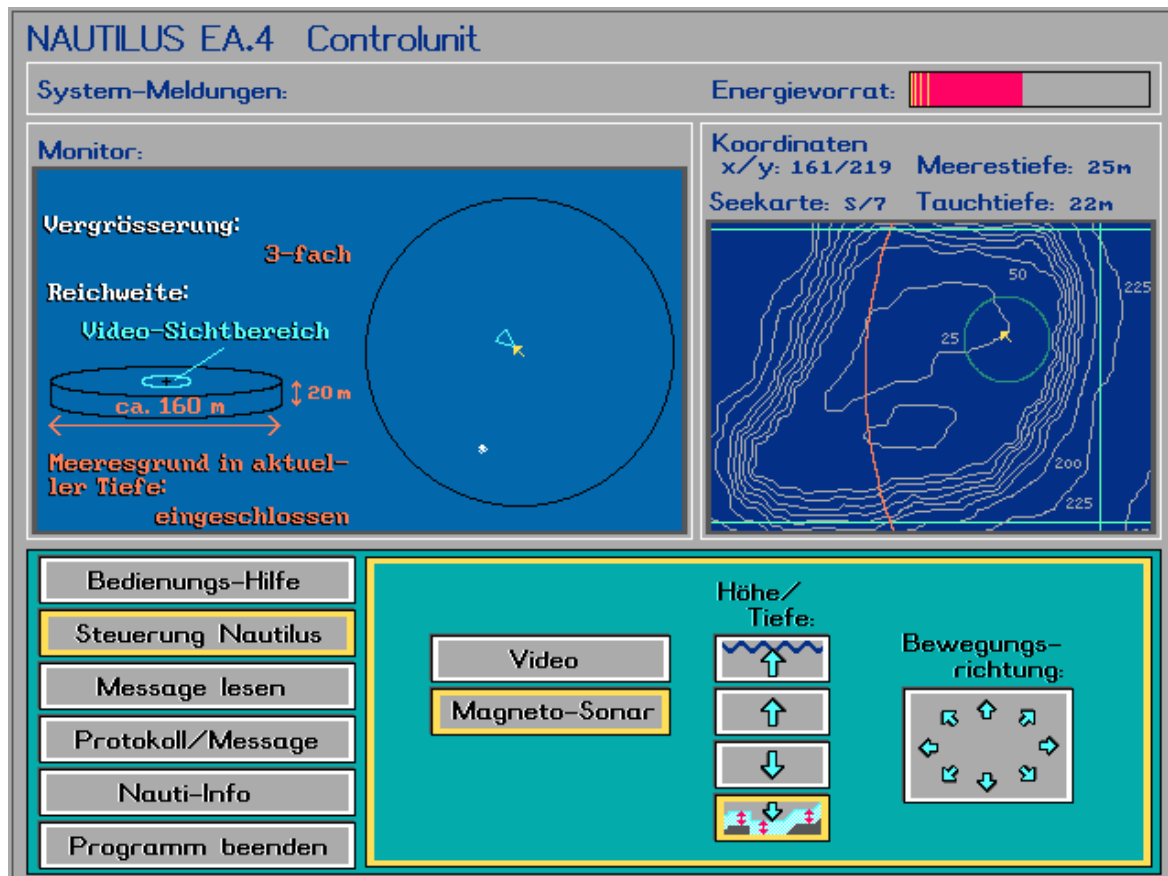


Abbildung 2: Steuerungscomputer für das ferngesteuerte Mini-U-Boot Nautilus. Unten links befindet sich das Hauptmenü, mit dem die verschiedenen Funktionseinheiten aufgerufen werden können, rechts davon wird das entsprechende Submenü eingeblendet. Auf dem Monitor oben links erscheint das Ergebnis des letzten Knopfdruckes, das heisst die zuletzt angeforderte Information (in diesem Beispiel das Ergebnis des Magneto-Sonar-Einsatzes, der weisse Punkt zeigt Metallvorkommen an). Der Abschnitt oben rechts enthält Angaben zum genauen Aufenthaltsort von Nautilus sowie einen Ausschnitt aus der Seekarte, der mit aktuellen Angaben ergänzt wird (der gelbe Pfeil symbolisiert Nautilus, die Pfeilspitze zeigt die Fahrtrichtung an, der grüne Kreis die ungefähre Reichweite des Sonargerätes etc.).

3.1.2 Darstellung des Spielverlaufs

Die Informationen über den simulierten Realitätsausschnitt sind Bestandteil der *Einführung* in das Szenario. Der Proband erfährt zudem, dass er sich erst anhand einer Trainingsaufgabe mit der Bedienung vertraut machen könne, um danach mit der eigentlichen Erzsuche zu starten. Die *Trainingsaufgabe* besteht darin, in einem eng umgrenzten Gebiet, dem Ausschnitt T/9 der Seekarte (vgl. Anhang B), über ein bereits bekanntes Erzvorkommen Informationen einzuholen und diese zu protokollieren. Um diese Aufgabe erfüllen zu können, werden alle Instrumente in typischer

Weise benötigt. Der Proband sollte durch Nachlesen in der Bedienungshilfe oder durch Ausprobieren herausfinden, wie er Nautilus fernsteuern kann, und insbesondere wie er das Sonargerät und die Videokamera anwenden muss, wie das Kommunikationssystem funktioniert und wo sich für die Problemlösung relevante Informationen befinden.

Es ist vorgesehen, dass zehn Minuten ausreichen, um die gewünschten Informationen über das bekannte Metallvorkommen in einem Protokoll festzuhalten. Erstellt der Proband innerhalb der ersten zehn Minuten nach Trainingsbeginn kein Protokoll, so erinnert ihn die Trainingsleitung an den noch ausstehenden Protokolleintrag (*Meldung T1*), liegt auch nach 15 Minuten noch kein Protokoll vor, so wird er definitiv zur Beendigung des Trainings mittels Protokolleintrag aufgefordert (*Meldung T2*). Sobald der Proband das *erste Protokoll* geschrieben hat, fordert ihn die Trainingsleitung auf, das Training zu beenden, zum Mutterschiff zurückzufahren und via Nachricht an die Technik das Auftanken von Nautilus anzufordern (*Meldung T3*). Der Auftankvorgang wird eingeleitet, sobald der Proband eine Nachricht an die Technik verschickt hat und Nautilus korrekt platziert ist (direkt am Forschungsschiff und an der Wasseroberfläche). Ist Nautilus falsch platziert, wird der Proband durch eine weitere Nachricht darauf hingewiesen, die Position von Nautilus zu korrigieren (*Meldung T4*).

Der *Auftankvorgang* dauert ca. drei Minuten und kann nicht unterbrochen werden. Die Steuerungseinheit für das Mini-U-Boot ist während dieser Zeit deaktiviert, die anderen Funktionseinheiten stehen dem Probanden jedoch zur Verfügung.

Nach dem Auftankvorgang beginnt die *Testphase*: Der Proband soll während etwa einer Stunde (so lange reiche der Energievorrat von Nautilus) im übrigen Forschungsgebiet nach unbekanntem Erzvorkommen suchen. Doch schon wenige Sekunden nach Beginn der Testphase erhält er von der Einsatzleitung die Nachricht über ein vermutlich abgestürztes Passagierflugzeug und wird aufgefordert, sich mit Nautilus unverzüglich an dieser Suche zu beteiligen (*Notfallmeldung 1*). Die Mittel und Möglichkeiten sind die gleichen wie bei der Erzsuche, nur steht der Proband bei dieser Aufgabe unter einem viel grösseren Zeitdruck: Das in Frage kommende Gebiet ist so gross, dass es nicht in der vorgesehenen Bearbeitungszeit von ca. einer

Stunde systematisch abgesucht werden kann, zudem sinkt mit zunehmender Suchzeit die Wahrscheinlichkeit, noch Überlebende zu finden. Der Proband muss Prioritäten setzen und eine Suchstrategie entwickeln. Die Meldung von der Einsatzleitung enthält allerdings einen *Hinweis auf eine mögliche sinnvolle Einschränkung des Suchgebietes*: Es wird erwähnt, dass reale Rettungschancen nur bis zu einer Meerestiefe von 50 Meter bestünden. Im Einsatzgebiet von Nautilus erfüllen nur zwei Bereiche dieses Kriterium, der Bereich S/7 und der Bereich R/9 der Seekarte (vgl. Anhang B).

Die *ideale Suchstrategie* besteht darin, dass der Proband seine Suche in einem dieser beiden Gebiete beginnt. Dabei sollte er den Meeresgrund des gewählten Gebietes systematisch, effizient und vollständig mit dem Sonargerät von Nautilus nach Metallvorkommen absuchen, so dass kein „Flecken“ ungeprüft bleibt. Da das Gerät eine Reichweite von 20 Metern hat, darf Nautilus nur eine maximale Distanz von 20 Metern zum Meeresgrund haben, damit (gesunkenes) Metall vom Sonargerät angezeigt wird. Für eine effiziente und flächendeckende Suche ist es notwendig, die Sonareinsätze in einem bestimmten Abstand zueinander zu machen. Zeigt das Sonargerät *keinen Metallfund* an, so sollte die nächste Sonaraufnahme möglichst viel neues und möglichst wenig bereits gescanntes Gebiet abdecken, das heisst Nautilus sollte um ca. einen halben bis ganzen Radius des Sonargerätes verschoben werden. Sobald das Sonargerät *einen Metallfund anzeigt*, so sollte der Proband diesen mit Hilfe der Videokamera näher untersuchen, um entscheiden zu können, ob es sich bei diesem Fund um das verunglückte Flugzeug handelt oder nicht. Dazu muss er Nautilus so manövrieren, dass sich der Metallfund unmittelbar in dem (im Sonarschirm eingezeichneten) Sichtbereich der Videokamera befindet (vgl. Abbildung 2, Seite 53). Während des Fahrens mit Nautilus sieht der Proband jedoch nichts und kann auch kein Metall wahrnehmen, das heisst für eine korrekte Ausrichtung von Nautilus beziehungsweise der Videokamera sind in der Regel mehrere kleinste Positionskorrekturen mit anschliessender Sonaraufnahme notwendig. Der Sichtbereich der Videokamera ist sehr gering (ca. vier Meter), so dass auch nach einer nur minimalen Verschiebung von Nautilus die Ausrichtung erneut mittels Sonargerät überprüft werden muss. Aufgrund der geringen Sichtweite der Kamera ist es auch

wenig ratsam, den Meeresgrund nur mit der Videokamera abzusuchen (d.h. ohne vorhergehenden Sonareinsatz): Um das ganze Gebiet vollständig zu untersuchen, müssten extrem viele Videobilder gemacht werden. Zudem benötigt das Übermitteln der Bilder viel Energie (pro Tankfüllung können ca. 100 Videobilder übermittelt werden).

All diese Informationen findet der Proband in der Bedienungshilfe und in der Datenbank Nauti-Info, oder teilweise auch durch Ausprobieren. Bei „falschen“ Sonar- und Videoeinsätzen (d.h. mehr als 20 Meter Distanz zum Meeresboden beim Sonar, mehr als drei Meter beim Video) erscheint auf dem Monitor eine kurze Information darüber, dass der Meeresgrund bei diesem Bild nicht eingeschlossen war. Zudem ist bei jeder Sonaraufnahme der Videosichtbereich auf dem Sonarschirm eingetragen.

Der weitere *Spielverlauf* gestaltet sich je nach *Suchstrategie* des Probanden unterschiedlich. Zum einen gibt es Systemereignisse, die nur unter bestimmten Bedingungen und bei einem bestimmten Verhalten auftreten. Zum anderen können Systemmeldungen, die für alle die gleiche Information enthalten, je nach Suchstrategie und Überlegungen des Probanden unterschiedliche Auswirkungen haben.

Macht ein Proband sechs bis neun Minuten nach Beginn der Testphase in einem der beiden sinnvollen Suchgebiete S/7 oder R/9 Videobilder vom Meeresgrund, so sieht er darauf ein *Gepäckstück (G)*, das den Probanden in seiner Suchstrategie bestätigen soll. In allen anderen Fällen (kein Video, Video nicht auf Meeresgrund, Suche in anderen Gebieten) wird kein Gepäckstück eingeblendet und der Proband verbleibt in Bezug auf die Richtigkeit seiner Überlegungen weiterhin im Dunkeln. Drei Minuten nach Sichtung des Gepäckstückes (oder neun Minuten nach Beginn der Testphase, falls kein Gepäckstück eingeblendet wurde) erfolgt eine Anfrage der Tauchergruppe, ob deren Einsatz gebraucht werde (*Meldung T*). Antwortet der Proband der Tauchergruppe, so dankt diese durch eine weitere Meldung und verspricht, sich bei Neuigkeiten wieder zu melden (*Meldung A*).

Im Abstand von fünf bis acht Minuten gehen anschliessend weitere Meldungen mit den neuesten Informationen über den möglichen Absturzort ein. Die *Meldungen 2 und 3* beinhalten beide (mit unterschiedlicher Deutlichkeit) den Hinweis, dass besser im rechten 50-Meter-Bereich (R/9) gesucht werden sollte. Beide Meldungen können,

je nach Suchstrategie des Probanden, ihn entweder in seiner Strategie bestätigen oder ihm nahe legen, das Gebiet zu wechseln. Für alle Probanden potentiell die gleiche Bedeutung hat die *Ort sensitive Meldung 4*: In Abhängigkeit vom gegenwärtigen Aufenthaltsort von Nautilus wird auf den jeweils gegenüber liegenden Bereich als sinnvoll für eine Suche hingedeutet, das heisst S/7, falls sich Nautilus rechts der Mitte, R/9 falls sich Nautilus links der Mitte befindet. Damit soll erreicht werden, dass mindestens ein Gebietswechsel vollzogen wird.

Den Abschluss der Bearbeitung bildet das Auffinden des Flugzeugwrackes. Probanden, die zwischen fünf und elf Minuten nach Erhalt der vierten Meldung irgendwo ein Videobild vom Meeresgrund machen, sehen darauf einen Teil des *Flugzeuges (F)*. Das bedeutet: Unter einer bestimmten Bedingung „finden“ alle Probanden, unabhängig von ihrer bisherigen Suchstrategie, das Flugzeug am Ort ihrer momentanen Suche. Das Erfolgserlebnis bleibt nicht nur den „guten“ Problemlösern vorbehalten. Wird nach Sichtung des Flugzeuges ein Protokoll erstellt oder eine Meldung verschickt, so gehen zwei weitere Meldungen der Einsatzleitung ein, mit Informationen über den Erfolg der Suche und bestem Dank an Nautilus (*Meldung 6 und 7*). Erstellt der Proband weder ein Protokoll noch eine Meldung, so erhält er die gleiche Meldung, die auch alle Probanden erhalten, die das Flugzeug nicht gefunden haben: die Nachricht, dass das Flugzeug von einem grossen Fischfänger (und Dank der Mithilfe von Nautilus) gefunden werden konnte und dass die Rettungschancen viel versprechend sind (*Meldung 5*).

Die *Spieldauer* variiert je nach Spielverhalten beträchtlich. Nach Angabe der Testautoren (Hirsig & de With, 2001) sind 75 bis 90 Minuten Bearbeitungszeit vorgesehen, wovon (gemäss der Instruktion) die ersten 15 Minuten zum Absolvieren der Trainingsaufgabe zur Verfügung stehen. Diese Vorgaben werden allerdings von vielen Probanden nicht erreicht: In der zur Verfügung stehenden Stichprobe von 230 Personen variieren die Werte für die *Gesamtspieldauer* zwischen 63 und 151 Minuten ($P_{25} = 77$, $M_d = 85$, $P_{75} = 94$), für die *Trainingsaufgabe* zwischen 12 und 64 (!) Minuten ($P_{25} = 20$, $M_d = 24$, $P_{75} = 31$) und für die *Testphase* (inkl. Auftankvorgang) zwischen 41 und 115 Minuten ($P_{25} = 54$, $M_d = 59$, $P_{75} = 65$).

3.1.3 Beschreibung des bestehenden Auswertungsansatzes

Die Konzeption von ATLANTIS macht deutlich, dass nicht der Zielzustand interessiert (da alle Probanden das Flugzeug „finden“ können), sondern der Weg zum Ziel, das heisst die Vorgehensweise, das Verhalten der Probanden. Klassische Problemlösegütemasse in Form von Steuerungsleistungen sind deshalb nicht definierbar. Die bestehende Auswertung fokussiert stattdessen das detaillierte Aufzeigen der Effizienz und Veränderbarkeit der Suchstrategie und der Aktivitäten des Probanden. Sie umfasst zwei Teile: eine Graphik über den Suchweg und die Suchstrategie und einen umfangreichen Protokollbogen mit diversen Kennzahlen zu ausgewählten Aspekten des Bearbeitungsprozesses. Diese beiden Bestandteile der Auswertung werden im Folgenden näher erläutert (Kapitel 3.1.3.1 und 3.1.3.2). Anschliessend folgen einige Erläuterungen zur Anwendung dieser Auswertung im Rahmen der Personalauswahl und -entwicklung des erwähnten Unternehmens (Kapitel 3.1.3.3).

3.1.3.1 Suchweg und Suchstrategie

Mit der Graphik über den Suchweg und die Suchstrategie (vgl. Abbildung 3) sollen wichtige Aspekte des Bearbeitungsprozesses visualisiert werden. Die Graphik besteht aus zwei Darstellungen, die den zeitlichen Ablauf *einzelner* Arbeitsschritte des Probanden nach unterschiedlichen Gesichtspunkten beschreiben. Zusammen sagen sie etwas darüber aus, wie gut der Proband die relevanten Aspekte der Problemstellung verstanden hat. In beiden Darstellungen sind der Aktionskreis von Nautilus, der Startpunkt und die beiden (während der Testphase) sinnvollen Suchbereiche eingezeichnet, die untere Darstellung zeigt zudem alle möglichen Erzvorkommen (Sonar-Ziele).

Beide Darstellungen zeigen den Suchweg des Probanden in Form einer Linie, deren Farbe den Steuerungsmodus charakterisiert (an Oberfläche, handgesteuert, automatische Abstandseinstellung). Die gestrichelte Linie in der oberen Darstellung markiert zudem die während dem Training zurückgelegte Strecke. In beiden Darstellungen sind auch alle Sonar- und Videoeinsätze eingetragen, mit unterschiedlichem Farbton

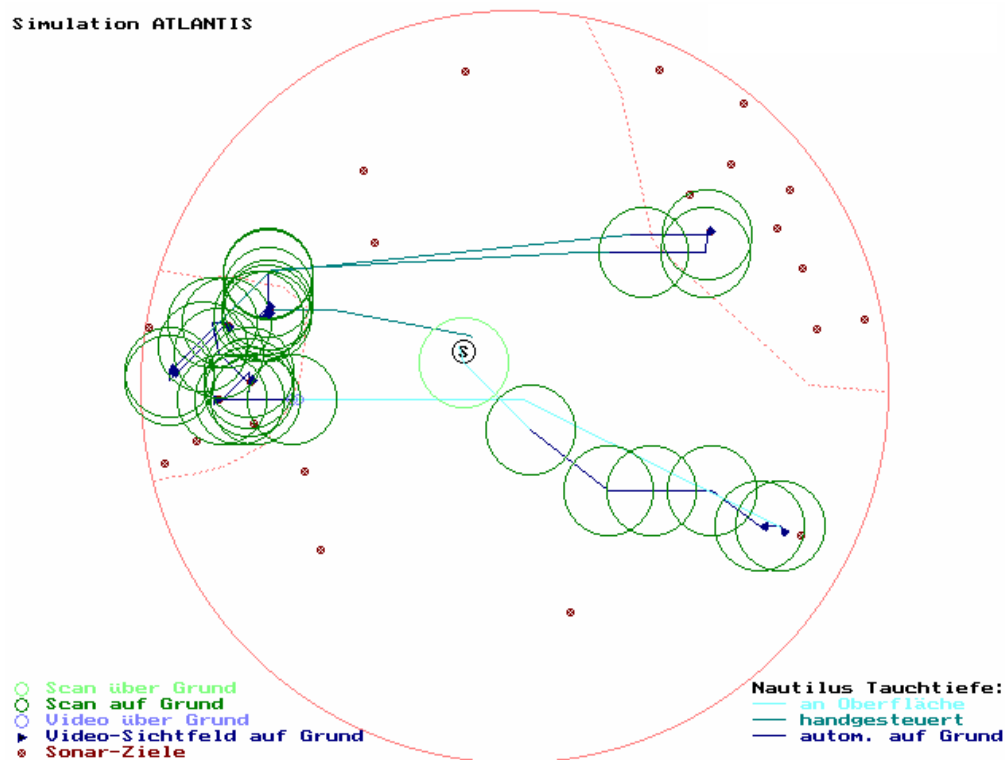
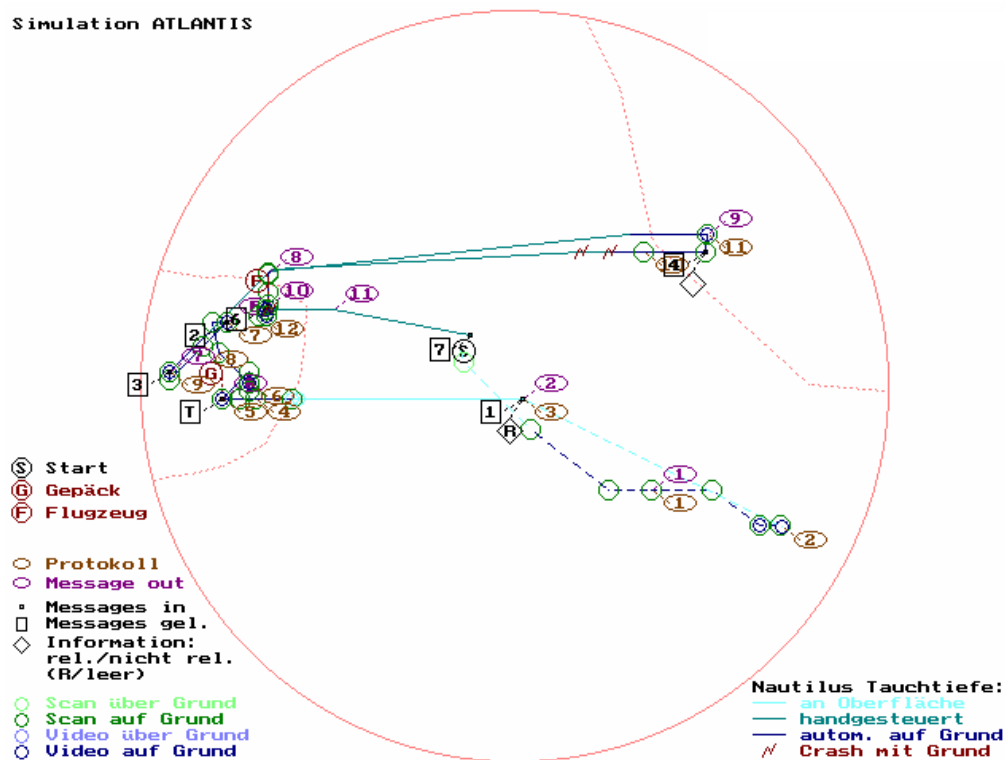


Abbildung 3: Graphische Darstellung des Suchwegs und der Suchstrategie bei der Bearbeitung von ATLANTIS. Die Graphik wird im Text näher erläutert.

für einen Einsatz über (hell) und auf (dunkel) dem Meeresgrund („über Grund“ bedeutet mehr als 20 (beim Sonar) bzw. drei (beim Video) Meter über Grund).

Die obere Darstellung fokussiert primär den *Suchweg* und den allgemeinen *Spielverlauf*. In den Suchweg sind auch alle von der Einsatzleitung *eintreffenden Nachrichten* (schwarze Nummern in Quadraten, für Abkürzungen vgl. Kapitel 3.1.2), die vom Probanden erstellten *Protokolle und Meldungen* (braune bzw. violette Nummern in Ellipsen) und die von ihm *abgerufenen Informationen* (schwarze Nummern in Rauten, R für relevante Information) sowie weitere *Systemmeldungen* (Gepäck, Flugzeug, Crash) eingetragen. Dieser Darstellung kann entnommen werden, wie der Proband auf die sich ständig verändernde Problemsituation reagiert.

Die untere Darstellung fokussiert mehr die *Suchstrategie*. Die Kreise beziehungsweise deren Überlappungen zeigen massstabsgetreu, wie flächendeckend und damit effizient, aber auch wie zielorientiert der Proband das Gebiet mit dem Sonargerät absucht. Die Kreissegmente (Dreiecke) kennzeichnen den Einsatz, die Richtung und die Sichtweite der Videokamera. Dieser Darstellung kann entnommen werden, inwiefern der Proband die Suchinstrumente korrekt und im Sinne der Aufgabenstellung einsetzt.

3.1.3.2 Protokollbogen mit Leistungskennwerten

Der Protokollbogen umfasst ca. acht bis zehn Seiten und beinhaltet detaillierte Angaben zur *Steuerung* von Nautilus, zur *Kommunikation* und zur *Informationssuche*. Die Angaben erfolgen in Form von Kennwerten, die entweder auf der Auszählung einzelner Aktionen respektive der Summierung von Bearbeitungszeiten beruhen (wie z.B. „Anzahl Magneto-Sonareinsätze“ oder „Lesezeit im Hilfesystem“), oder durch das Auszählen ganzer Handlungssequenzen zustande kommen (wie z.B. „Anzahl optimal versetzter Scans“ oder „Anzahl Videoeinsätze mit Metall im Videosichtbereich“). Ein weiterer Bestandteil des Protokollbogens ist die exakte Beschreibung und Auflistung ausgewählter Aktionen. So ist beispielsweise ersichtlich, welche

einzelnen Seiten der Bedienungshilfe oder der Datenbank Nauti-Info der Proband zu welchem Zeitpunkt aufgerufen hat oder wann er zum ersten Mal die Steuerungsautomatik eingeschaltet hat. Ebenfalls wiedergegeben sind die Inhalte der vom Probanden verfassten Protokolle und Meldungen.

Der Protokollbogen besteht aus fünf Abschnitten, die zum Teil redundante Angaben enthalten. Die Abschnitte umfassen:

1. Eine kurze *Übersicht* über die gesamte Bearbeitung, mit Angaben über die Dauer des Trainings und der Testphase, den relativen Anteil der Lesedauer sowie Angaben über die verschiedenen Sonar- und Videoeinsätze.
2. Den zeitlichen Ablauf der *Kommunikation*, der aufzeigt, zu welchen Zeitpunkten die eingegangenen Meldungen (wie oft und mit welcher zeitlicher Verzögerung) gelesen wurden, und zu welchen Zeitpunkten der Proband was geschrieben hat (Inhalte aller erstellten Texteinheiten).
3. *Kennwerte zur Trainingsaufgabe*, mit detaillierten Angaben über das Aufrufen einzelner Seiten der Bedienungshilfe und der Datenbank Nauti-Info, Angaben über den zeitlichen Ablauf von „Meldungen lesen“ und „Texteinheiten schreiben“ (jedoch im Unterschied zum Punkt 2 ohne Textinhalt), Angaben über den erstmaligen Einsatz der Steuerungsmodi von Nautilus sowie Angaben über die verschiedenen Sonar- und Videoeinsätze (Zeitpunkt, Anzahl).
4. *Kennwerte zur Flugzeugsuche*, mit detaillierten Angaben analog 3.
5. *Ausgewählte Kennwerte* zur problembezogenen Aktivität und Effizienz, zur Problemerkennung und -lösung sowie zum Kommunikationsfluss. Diese 13 Kennwerte fassen die wichtigsten Aspekte der Problembearbeitung zusammen und werden als Leistungsmasse verstanden. Sie sollen dem Diagnostiker eine rasche, erste Einschätzung der Problemlösefähigkeit ermöglichen.

Eine detaillierte Beschreibung und Erläuterung des gesamten Protokollbogens soll an dieser Stelle nicht erfolgen. Die meisten der aufgeführten Kennwerte, Zahlen und Angaben sind, wie dem Beispiel-Protokollbogen in Anhang B entnommen werden kann, selbst erklärend. Sie beruhen, wie bereits erläutert, auf der Auszählung von Aktionen und Handlungssequenzen oder der Aufsummierung von Bearbeitungszeiten. Kennwerte, deren Berechnung weniger eindeutig ist, sind in Tabelle 2 beschrieben. Es handelt sich um die auf der letzten Seite des Protokollbogens dargestellten 13 Kennwerte beziehungsweise Leistungsmasse, die für die weitere Arbeit von Bedeutung sein werden. Alle 13 Kennwerte basieren auf relativen Häufigkeiten und werden auf einer Skala von 0-100 abgebildet. Im Sinne der in der Literatur verwendeten Terminologie (vgl. Kapitel 2.1) handelt es sich um *Verhaltensmasse* oder

Verhaltensskalen. Im Folgenden wird jedoch ausschliesslich der Begriff (*szenariospezifische*) *Leistungskennwerte* verwendet, damit die Abgrenzung gegenüber den durch das neue Auswertungsmodell eingeschätzten Massen deutlich wird. Bei diesen handelt es sich ja ebenfalls um Verhaltensmasse (mikroverhaltensbasierte Prozessmasse und Prozessprofile).

3.1.3.3 Anwendung der Auswertung in der Praxis

Die vorangegangenen beiden Kapitel sollten zeigen, wie detailliert und umfangreich der bestehende Auswertungsansatz von ATLANTIS ist. Wird die gesamte zur Verfügung stehende Information genutzt, so kann ein umfassendes, ganzheitliches Bild des Bearbeitungsprozesses entstehen. Analog zu anderen Problemlöseszenarien liegen jedoch auch für ATLANTIS keine Regeln zur Interpretation und Integration der verschiedenen Daten vor. Die korrekte Anwendung der Auswertung erfordert nicht nur viel Zeit, sondern auch eine gute Kenntnis des gesamten Szenarios.

Immerhin ermöglichen die 13 Leistungskennwerte, die auf der letzten Seite des Beurteilungsbogens die wichtigsten Aspekte des Problemlöseverhaltens zusammenfassen, eine rasche, erste Einschätzung der Problemlöseleistung. Für diese liegen auch eine Art *Normwerte* vor, die dem Diagnostiker helfen sollen, die erhaltenen Werte zu interpretieren. Gute Problemlöser sollten demnach in den 13 Leistungskennwerten mindestens die folgenden Punkte erreichen (zur Erinnerung: Die Skalen reichen von 0-100; für Abkürzungen der Kennwerte vgl. Tabelle 2):

- <i>p_metall</i>	40	- <i>p_ow_m</i>	90	- <i>p_s_mg</i>	95	- <i>p_m_p_nv</i>	100
- <i>p_met_s7</i>	60	- <i>p_ow_i</i>	100	- <i>p_v_mg</i>	90	- <i>p_mess15</i>	95
- <i>p_met_r9</i>	30	- <i>p_scan_v</i>	80	- <i>p_s_mg79</i>	95	- <i>p_iseit</i>	100
				- <i>p_v_mv79</i>	80		

Diese Grenzwerte oder Richtwerte entstanden durch die Beschreibung einer ersten Probandengruppe, die im Rahmen der Personalselektion des erwähnten Unternehmens mit ATLANTIS getestet wurde. Die Grenzwerte liegen wenige Punkte über dem dritten Quartilwert und beschreiben damit die Werte, die von weniger als 25% der Probanden erreicht oder überschritten werden.

Tabelle 2: Übersicht über die 13 szenariospezifischen Leistungskennwerte (Verhaltensmasse) von ATLANTIS.

Bezeichnung im Protokollbogen	Berechnungsgrundlage und inhaltliche Bedeutung	Abkürzung
<i>Kennwerte zur problembezogenen Aktivität und Effizienz (1-3 beschreiben mehr das Agieren, 4-6 die Flexibilität)</i>		
1) x % der potentiellen Fundorte in beiden Suchgebieten wurden gescannt	Prozentualer Anteil der Fläche in den sinnvollen Suchgebieten S/7 und R/9 (= 19 Metallvorkommen), die der Proband mit dem Sonargerät abgesucht hat. <i>Wie flächendeckend hat der Proband in den beiden sinnvollen Suchgebieten gescannt?</i>	p_metall
2) x % der potentiellen Fundorte im linken Suchgebiet (S/7) wurden gescannt	Prozentualer Anteil der potentiellen Fläche in S/7 (9 Metallvorkommen). <i>Wie flächendeckend hat der Proband im linken sinnvollen Suchgebiet gescannt?</i>	p_met_s7
3) x % der potentiellen Fundorte im rechten Suchgebiet (R/9) wurden gescannt	Prozentualer Anteil der potentiellen Fläche in R/9 (10 Metallvorkommen). <i>Wie flächendeckend hat der Proband im rechten sinnvollen Suchgebiet gescannt?</i>	p_met_r9
4) x % der Bearbeitungszeit für eingegangene Informationen wurden effizient genutzt	Lesezeit für nicht wiederholte Meldungs-Seiten in Relation zur Lesezeit aller Meldungs-Seiten. <i>Hat der Proband die einkommenden Meldungen effizient, das heisst ohne Wiederholungen gelesen?</i>	p_ow_m
5) x % der Bearbeitungszeit für abgerufene Informationen wurden effizient genutzt	Lesezeit für nicht wiederholte Nauti-Info-Seiten in Relation zur Lesezeit aller Nauti-Info-Seiten. <i>Hat der Proband die abgerufenen Informationsseiten effizient, das heisst ohne Wiederholungen gelesen?</i>	p_ow_i
6) x % der Verschiebungen der Suchpunkte wurden effizient gestaltet	Prozentualer Anteil an Sonar-Verschiebungen x, mit „ $x \leq 1$ Radius“ nach Scans mit Metall und „ $1 \text{ Radius} \leq x \leq 2 \text{ Radien}$ “ nach Scans ohne Metall. <i>Wie effizient hat der Proband seine Suche (d.h. die Sonareinsätze) gestaltet?</i>	p_scan_v

Tabelle 2 (Fortsetzung): Übersicht über die 13 szenariospezifischen Leistungskennwerte (Verhaltensmasse) von ATLANTIS.

Bezeichnung im Protokollbogen	Berechnungsgrundlage und inhaltliche Bedeutung	Abkürzung
<i>Kennwerte zur Problemerkennung und -lösung</i>		
7) x % der Suchaktivitäten wurden in sinnvoller Tiefe (auf Grund) durchgeführt	Prozentualer Anteil an Sonareinsätzen, die der Proband auf Grund (≤ 20 Meter Distanz zu Meeresgrund) durchgeführt hat. <i>Hat der Proband mehrheitlich sinnvolle Sonareinsätze gemacht?</i>	p_s_mg
8) x % der Prüfsaktivitäten wurden in sinnvoller Tiefe (auf Grund) durchgeführt	Prozentualer Anteil an Videoeinsätzen, die der Proband auf Grund (≤ 3 Meter Distanz zu Meeresgrund) durchgeführt hat. <i>Hat der Proband mehrheitlich sinnvolle Videoeinsätze gemacht?</i>	p_v_mg
9) x % der Suchaktivitäten wurden in sinnvoller Tiefe und in sinnvollen Gebieten durchgeführt	Prozentualer Anteil an Sonareinsätzen, die der Proband auf Grund und in den beiden sinnvollen Suchgebieten (S/7 und R/9) durchgeführt hat. <i>Hat der Proband mehrheitlich sinnvolle Sonareinsätze in sinnvollen Gebieten gemacht?</i>	p_s_mg79
10) x % der Prüfsaktivitäten bezogen sich genau auf potentielle Funde in sinnvollen Gebieten	Prozentualer Anteil an Videoeinsätzen, die der Proband auf Grund und in den beiden sinnvollen Suchgebieten (S/7 und R/9) und nach gezieltem Manövrieren (so dass sich Metall genau im Sichtbereich befand) durchgeführt hat. <i>Hat der Proband mehrheitlich korrekte Videoeinsätze in sinnvollen Gebieten gemacht?</i>	p_v_mv79
<i>Kennwerte zum Kommunikationsfluss</i>		
11) x % der Informationen wurden effizient aus-/ weitergegeben	Prozentualer Anteil an (vom Probanden erstellten) Texteinheiten, die er nicht verworfen hat und die nicht leer waren. <i>Hat der Proband seine eröffneten Protokolle und Meldungen gespeichert bzw. verschickt?</i>	p_m_p_nv
12) x % der ankommenden Informationen wurden effizient aufgenommen/abgerufen	Prozentualer Anteil an eingegangenen Meldungen, die der Proband in weniger als 15 Sekunden nach ihrem Eingang gelesen hat. <i>Hat der Proband die einkommenden Meldungen rasch gelesen?</i>	p_mess15
13) x % der ankommenden Informationen wurden vollständig abgerufen	Prozentualer Anteil an eingegangenen Meldungen, die der Proband vollständig (d.h. alle Seiten) gelesen hat. <i>Hat der Proband die einkommenden Meldungen vollständig gelesen?</i>	p_iseit

Viele Probanden erreichen jedoch nur in einigen Kennwerten die vorgegebenen Grenzwerte, in anderen nicht. Wie solche inkonsistenten Ergebnisse zu interpretieren sind, bleibt im Ermessen des jeweiligen Diagnostikers. Die Testautoren weisen allerdings darauf hin, dass „die *Kennwerte der Problemerkennung und -lösung* (p_s_mg, p_v_mg, p_s_mv79, p_v_mv79) das K.O.-Kriterium darstellen“⁴. Ausserdem sollte ein guter Problemlöser auch in den *Kennwerten des Agierens* (p_metall, p_met_s7, p_met_r9) die vorgegebenen Grenzwerte erreichen, wobei der Kennwert p_met_r9 im Sinne von „Flexibilität“ und weniger als reines Leistungsmass zu interpretieren ist (der Kennwert zeigt, ob auch ein Wechsel der Suchgebiete stattgefunden hat). In den anderen sechs Kennwerten müssen die Grenzwerte nicht durchgehend erreicht sein, damit ein Proband noch als guter Problemlöser gilt.

Auch für einige weitere Kennzahlen des Protokollbogens liegen solche Richtwerte vor (z.B. Trainingsdauer nicht länger als 20 Minuten, mindestens fünf Sonareinsätze im Training, von denen mindestens 80% auf Grund sind etc.). Ob und wie die Diagnostiker diese Angaben verwenden, welche Bestandteile der Auswertung zur Einschätzung des Problemlöseverhaltens und Beurteilung der Probandenleistung sie benutzen, ist nicht bekannt. Das Unternehmen, das ATLANTIS seit einigen Jahren lizenziert hat, verwendet das Instrument im Rahmen von Assessment Centern für die Personalauswahl und -entwicklung (vorwiegend für die Besetzung von Führungs- und Kaderpositionen). ATLANTIS trägt damit nur als ein Baustein zum Gesamturteil bei. Über die weiteren eignungsdiagnostischen Verfahren, die je nach Fragestellung zur Beurteilung potentieller Kandidaten herangezogen werden, liegen leider keine Angaben vor.

⁴ Persönliche Mitteilung der Testautorin Dr. Astrid E. de With im August 2006.

3.2 Entwicklung des neuen Auswertungsmodells

Die folgenden Abschnitte erläutern die schrittweise Entwicklung des neuen Auswertungsmodells. In einem ersten Schritt wird – aufgrund inhaltlicher Überlegungen – ein Satz von Interpretationsregeln definiert, anhand dessen die einzelnen Aktionen der Probanden im Hinblick auf das Informationseröffnende und Adäquate Handeln beurteilt und bewertet werden können (Prozessorientierte Mikroverhaltensbeschreibung, Kapitel 3.2.1). In einem zweiten Schritt wird die zeitliche Entwicklung der beiden Konstrukte graphisch dargestellt und festgelegt, anhand welcher Kenngrößen die Zeitreihen formal beschrieben werden können (Kapitel 3.2.2). Als drittes wird – unter Einbezug formaler Klassifizierungsmethoden – die Ähnlichkeit der Zeitreihen beurteilt und ein Satz von Klassifikationsregeln definiert, der die Prozessprofile für das Informationseröffnende und Adäquate Handeln (und damit die Problemlösetypen, d.h. die Einordnung im Beurteilungsschema) beschreibt (Kapitel 3.2.3).

3.2.1 Prozessorientierte Mikroverhaltensbeschreibung

Die prozessorientierte Mikroverhaltensbeschreibung hat zum Ziel, jeden registrierbaren Handlungsschritt eines Probanden im aktuellen Handlungskontext zu beurteilen und zu bewerten. Als erstes stellt sich damit die Frage, welche Handlungsschritte von ATLANTIS automatisch registriert werden, das heisst wie die verfügbare Datenbasis aussieht (Kapitel 3.2.1.1). Darauf aufbauend kann überlegt werden, unter welchen Bedingungen die einzelnen Aktionen im Hinblick auf das Informationseröffnende und Adäquate Handeln bewertet werden sollen (Kapitel 3.2.1.2 bis 3.2.1.4).

3.2.1.1 Datenbasis und allgemeine Definitionen

Nach Abschluss der Bearbeitung von ATLANTIS liegt von jedem Probanden ein automatisch vom Rechner generiertes *Verhaltensprotokoll* vor. Dieses umfasst die zeit-synchrone Auflistung aller Arbeitsschritte des Probanden, das heisst alle automatisch

registrierbaren *Aktionen* des Probanden (Mausklicks). Zu jeder Aktion wird auch der *aktuelle Systemzustand* gespeichert: die Systemzeit (Zeitdauer seit Spielbeginn), Angaben über die Position von Nautilus (Koordinaten, Seekartenausschnitt, Tauch- und Meerestiefe) und Angaben über den Zustand beziehungsweise die Umgebung von Nautilus (Energievorrat, Ausrichtung der Videokamera, potentielle Metallziele). Das Verhaltensprotokoll enthält auch *alle Systemmeldungen*, das heisst Nachrichten, die der Proband von anderen Sucheinheiten erhält, sowie Informationen vom oder über das System (z.B. die Information über eine unmittelbar bevorstehende Kollision mit dem Meeresgrund). Zu den Systemmeldungen gehören auch Angaben über die Ergebnisse von Sonar- und Videoeinsätzen.

Die zeitliche Abfolge der Arbeitsschritte einer Versuchsperson, die dazugehörigen Systemmeldungen und Angaben über den Systemzustand (sowie daraus berechneten weiteren Angaben wie die zurückgelegte Distanz, der Abstand zum Meeresgrund etc.) können in tabellarischer Form aufbereitet werden. Für eine durchschnittliche Bearbeitung des Szenarios liegen 300 bis 500 Datenzeilen vor, sowie die *Inhalte der vom Probanden verfassten Protokolle und Nachrichten*, die separat, nach Ausgabe aller Datenzeilen, dargestellt werden. Anhang C zeigt einen Ausschnitt aus einem auf diese Weise aufbereiteten Verhaltensprotokoll (ohne Inhalt der Protokolle und Nachrichten). In der letzten Spalte finden sich – in Textform – die Systemmeldungen und die Aktionen des Probanden, die die Grundlage bilden für die prozessorientierte Mikroverhaltensbeschreibung.

Diese *Textausgaben* lassen sich grob in drei Handlungskategorien einteilen, die unterschiedliche Funktionseinheiten des Steuerungscomputers Nautilus EA.4 betreffen (vgl. dazu Kapitel 3.1.1):

- Die Kategorie *Lesen* betrifft die Bedienungshilfe, die Datenbank Nauti-Info sowie die Funktion „Message lesen“ des Kommunikationssystems.
- Die Kategorie *Schreiben* betrifft die Funktion „Protokoll/Message“ des Kommunikationssystems.
- Die Kategorie *Suchen und Prüfen* betrifft die Steuerung von Nautilus, das heisst die Fortbewegung von Nautilus und den Einsatz von Sonar- und Videogerät.

Zur Handlungskategorie Lesen

Die Informationen in der Bedienungshilfe und in der Datenbank Nauti-Info sowie die eingegangenen Nachrichten können ein- oder mehrseitig sein, wobei der Proband jeweils von der auf dem Monitor angezeigten Seite schrittweise vorwärts und rückwärts blättern kann. Für jede einzelne aufgerufene Seite wird registriert, wann sie geöffnet wurde und wann die nächste Aktion erfolgte. Für eingegangene Nachrichten wird zudem der Sendezeitpunkt gespeichert.

Als Mass dafür, ob eine Seite als gelesen gilt oder nicht, wird für alle Seiten eine minimale Zeitdauer festgelegt, während der die Seite geöffnet sein muss. Diese Zeitdauer wird im Folgenden als *minimale Lesedauer* bezeichnet. Für die Bewertung der Lesedauer wird die Textlänge einer Seite und der Informationsgehalt berücksichtigt: Für kurze Seiten mit wenig Informationen (ca. fünf Zeilen, mehr als ein bis drei Sätze) beträgt die minimale Lesedauer fünf, für längere Seiten mit mehr Informationen zehn Sekunden. Sehr kurze Seiten mit nur ein bis zwei Zeilen gelten schon ab zwei Sekunden als gelesen (diese sind allerdings sehr selten). Wird kein Text angezeigt, dafür eine Übersicht über das verfügbare Menü, beträgt die minimale Lesedauer ebenfalls fünf Sekunden, wird zum Menü auch noch ein kurzer erklärender Text angezeigt, beträgt die minimale Lesedauer wiederum zehn Sekunden (für genauere Angaben vgl. Anhang C).

Das unmittelbar wiederholte Aufrufen der gleichen Seite wird als eine Aktion bewertet, indem die Zeitdauer der einzelnen Klicks zusammengezählt und die Summe mit der minimalen Lesedauer der Seite verglichen wird.

Für die Mikroverhaltensbeschreibung muss auch die *Relevanz* der abgerufenen Informationen für die Problemlösung berücksichtigt werden. Während alle Seiten der Bedienungshilfe sowie alle eingehenden Nachrichten von anderen Sucheinheiten relevante Informationen enthalten, gibt es in der Datenbank Nauti-Info nur insgesamt fünf Themen mit relevanten Informationen, alle anderen Themen sind entweder für die Problemlösung nicht unmittelbar von Bedeutung (= neutrale Informationen) oder haben rein unterhaltenden Wert. Welche Seiten als relevant, neutral oder unterhaltend gelten, kann Anhang C entnommen werden.

Zur Handlungskategorie Schreiben

Das System speichert, zu welchem Zeitpunkt der Proband ein Protokoll beziehungsweise eine Nachricht eröffnet hat, was er geschrieben hat, und zu welchem Zeitpunkt und an wen er die Nachricht verschickt (oder verworfen) hat beziehungsweise zu welchem Zeitpunkt er das Protokoll gespeichert (oder verworfen) hat. Ausgegeben wird auch, wann der Proband vom System zu einem Protokolleintrag aufgefordert wurde (Systemmeldung).

In der Aufgabenstellung wird das Schreiben von Protokollen und Meldungen explizit verlangt. Von daher gilt das Verfassen von Texteinheiten grundsätzlich als sinnvoll. Unter welchen Bedingungen Texteinheiten als nicht sinnvoll gelten, wird bei der Beschreibung des Adäquaten Handelns erläutert.

Zur Handlungskategorie Suchen und Prüfen

Das System speichert, wann und für wie lange der Proband Nautilus in welchem Modus bewegt hat, und wann eine Sonaraufnahme oder ein Videobild angefordert wurde. Für den Modus „Fahren in horizontaler Richtung“ wird nur registriert, wie lange Nautilus bewegt wurde und nicht, welche der acht Bewegungsrichtungen aktiviert worden war (Textausgabe „Move“ mit Zeitangabe). Die Systemmeldung „Crash“ zeigt an, wann der Proband über eine bevorstehende Kollision mit dem Meeresgrund informiert wurde. Die Systemmeldungen „Netz“ und „Wrack“ vor der Textausgabe „Video“ bedeuten, dass der Proband auf dem Videobild ein Gepäckstück (Netz) beziehungsweise einen Teil des havarierten Flugzeuges (Wrack) gesehen hat.

Für Sonar- und Videoeinsätze kann dem Verhaltensprotokoll – mit etwas Übung – das Resultat der angeforderten Bilder entnommen werden. Für *Sonareinsätze* sind die Textausgaben folgendermassen zu interpretieren:

- Textausgabe „Sonar: 0000000“ und > 20 Meter über Grund
= *Sonareinsatz ohne Meeresgrund („falscher“ Sonareinsatz)*
- Textausgabe „Sonar: 0000000“ und ≤ 20 Meter über Grund
= *Sonareinsatz mit Meeresgrund, aber ohne Metallfund*

- Textausgabe „Sonar: 0000000“ und ≤ 20 Meter über Grund und Angabe von Koordinaten unter Sonarziele
= *Sonareinsatz mit Meeresgrund, mit Metallfund im Scan-, aber nicht im Videosichtbereich*
- Textausgabe „Sonar: 01001000“
= *Sonareinsatz mit Meeresgrund, mit Metallfund im Videosichtbereich* (Der Aktionsradius des Sonargerätes ist insgesamt in acht gleich grosse Sektoren aufgeteilt, deren Reihenfolge im Uhrzeigersinn der Position der Nullen und Einsen in der Textausgabe entspricht. Eine Eins steht für Metallvorkommen im entsprechenden Sektor. Im Beispiel „Sonar: 01001000“ befindet sich Metall im Sektor 2 und 5)

In Kapitel 3.1.2 wurde bereits erläutert, wie Sonar- und Videogerät für eine effiziente Suche idealerweise einzusetzen sind (zuerst mit dem Sonargerät den Meeresgrund nach Metall absuchen, dann den Metallfund – durch exaktes Ausrichten von Nautilus, meist verbunden mit mehreren Sonareinsätzen – mit der Videokamera prüfen). Für die Mikroverhaltensbeschreibung gilt ein *Videoeinsatz* nur dann als *korrekt*, wenn die folgenden drei Bedingungen erfüllt sind:

- a) Distanz zum Meeresgrund ist kleiner als 4 Meter.
- b) Vor dem Videoeinsatz erfolgte ein Sonareinsatz mit Grund und Metall im Videosichtbereich (Sonar: 01001000), und zwischen Sonar- und Videoeinsatz erfolgte keine Ortsveränderung (Änderungen der Richtung von Nautilus oder der Tauchtiefe sind erlaubt, sofern a und c erfüllt sind).
- c) Nautilus ist so ausgerichtet, dass sich der Metallfund im Videosichtbereich befindet (d.h. die Richtung von Nautilus stimmt mit den Treffer-Sektoren überein; im Beispiel von oben müsste die Richtung von Nautilus auf Position 2 oder 5 sein).

Grundsätzlich wäre auch denkbar, die Bedingung b wegzulassen und alle Videoeinsätze mit Grund und Metall im Videosichtbereich positiv zu werten (wie dies im Übrigen bei der Berechnung der szenariospezifischen Leistungskennwerte gemacht wird, vgl. Kapitel 3.1.3.2). Damit würde berücksichtigt, dass jemand ohne Sonareinsätze auch nur mit der Videokamera suchen (und Treffer erzielen) kann. Allerdings gibt es nur sehr wenige Probanden, die die Videokamera nicht im Rahmen von Sonareinsätzen verwenden, und zudem sind deren Videoeinsätze meistens nicht auf Grund (d.h. Bedingung a ist nicht erfüllt).

Die drei Handlungskategorien decken die vom Programm registrierten Arbeitsschritte des Probanden vollständig ab. Offen bleibt, wann der Proband etwas anderes macht als lesen, schreiben und suchen/prüfen. So ist beispielsweise nicht beurteilbar,

ob der Proband eine geöffnete Seite der Bedienungshilfe tatsächlich liest oder ob er währenddessen Einträge auf der Seekarte vornimmt oder ob er einfach nur über das weitere Vorgehen nachdenkt. Feststellen lässt sich nur, wann und für wie lange der Proband eine bestimmte Funktion der Steuerungseinheit aktiviert hat. Wie bei allen Logfile-Daten handelt es sich um reine *Verhaltensdaten*, die keine direkte Erfassung kognitiver Prozesse erlauben. Auf die einem Handlungsschritt zugrunde liegenden kognitiven Prozesse oder auf dessen Bedeutung im jeweiligen Handlungskontext kann deshalb nur geschlossen werden. Eine Ausnahme bilden die vom Probanden erstellten Protokolle und Nachrichten, die Überlegungen des Probanden zum Stand der Suche und zum weiteren Vorgehen enthalten. Allerdings ist auch hier offen, inwiefern diese Angaben wirklich das Denken des Probanden abbilden und inwiefern sie nicht beispielsweise durch soziale Erwünschtheit verzerrt sind.

Wie die einzelnen Handlungsschritte nun im Hinblick auf das Informationseröffnende und Adäquate Handeln bewertet werden, ist Bestandteil der folgenden beiden Kapitel. Für beide Konstrukte werden einige grundlegende Regeln der Bewertung erläutert, der gesamte Satz an Interpretationsregeln ist Anhang C zu entnehmen.

3.2.1.2 Kriterien für das Informationseröffnende Handeln

Mit diesem Konstrukt wird erfasst, wie gut und effizient der Proband während der Problembearbeitung immer wieder neue, bezüglich der aktuellen Aufgabenstellung relevante Informationen sucht. Der Proband kann sich durch Nachschlagen in der Bedienungshilfe oder der Datenbank Nauti-Info problemrelevante Informationen verschaffen, beispielsweise Informationen über die korrekte Anwendung von Sonar- und Videogerät. Die von anderen Sucheinheiten eingegangenen Nachrichten enthalten zudem neueste Informationen über den aktuellen Stand der Suche und geben laufend Hinweise, wie die Suche sinnvoll eingeschränkt werden kann. Und durch den korrekten Einsatz von Sonar- und Videogerät gewinnt der Proband zunehmend mehr Informationen über den möglichen Absturzort des Flugzeuges (bzw. über Metallvorkommen während der Trainingsaufgabe).

Die Bewertung der einzelnen Aktionen erfolgt auf einer dreistufigen Skala, deren Ausprägungen das Ausmass potentieller Informationsaufnahme abbilden. Eine Aktion kann dem Probanden *viel* (= +1), *wenig* (= +0.5) oder *keine* (= 0) *neue und relevante Information* verschaffen. Um zu beurteilen, ob eine Information neu und relevant ist, muss der aktuelle Handlungskontext berücksichtigt werden, das heisst die aktuelle Aufgabenstellung und die „Vorgeschichte“ (was der Proband bisher gemacht hat).

Nur Informationen, die für die Problemlösung (mehr oder weniger) von Bedeutung sind, werden positiv bewertet (relevante und neutrale Informationen), das Nachschlagen von unwichtigen oder unterhaltenden Informationen gilt als nicht informationseröffnend (= 0). Ebenfalls keine Information verschaffen Aktionen wie Protokolle und Meldungen schreiben oder das Herumfahren von Nautilus. Sie sind zwar für die Problemlösung relevant, bringen aber für sich keine Erkenntnisse über die Situation. Bewertet werden damit nur die Aktionen der Handlungskategorie *Lesen* und *Suchen/Prüfen*.

Anhand der „Vorgeschichte“ wird beurteilt, ob die Information „neu“ ist: Das ständige Wiederholen von Aktionen verschafft keine neue Information. Beim *Lesen* wird davon ausgegangen, dass der Proband, wenn er eine Seite zum ersten Mal liest (d.h. während der minimalen Lesedauer geöffnet hat), diese Information auch aufnimmt (= +1). Liest er sie ein zweites Mal (vielleicht hat er etwas übersehen), so wird ihm noch ein wenig Informationsgewinn zugesprochen (= +0.5), ab dem dritten Mal jedoch verschafft das erneute Lesen der gleichen Seite keine neue Information mehr (= 0). Diese Definition trifft das tatsächliche Verhalten bestimmt nur zum Teil. So ist auch denkbar, dass der Proband beim ersten Öffnen der Seite diese gar nicht liest, sondern nachdenkt und erst beim erneuten Aufrufen der Seite die Information aufnimmt, oder er nimmt die Information überhaupt nicht auf, weil er sie nicht versteht. Die tatsächliche Aufnahme und Speicherung von Information kann anhand der zur Verfügung stehenden Verhaltensprotokolle nicht überprüft werden. Beurteilt wird nur die potentielle Informationsaufnahme, den theoretisch möglichen Informationszuwachs durch eine bestimmte Aktion. Ein weiterer Aspekt betrifft die Frage, ob das dadurch erworbene potentielle Wissen auch umgesetzt werden kann.

Diese Frage wird teilweise beantwortet durch die Berücksichtigung des Adäquaten Handelns (vgl. nachfolgendes Kapitel).

Die Bewertung von *Sonar- und Videoeinsätzen* erfolgt nach ähnlichen Überlegungen. Wiederholungen verschaffen keine neue Information, das heisst ein Einsatz am genau gleichen Ort wie ein früherer Sonar- beziehungsweise Videoeinsatz (gleiche Koordinaten, gleiche Tauchtiefe, gleiche Ausrichtung des Videosichtbereichs) wird als nicht informationseröffnend (= 0) bewertet. Nur durch Sonareinsätze auf Meeresgrund und korrekte Videoeinsätze verschafft sich der Proband relevante Informationen. Alle Videoeinsätze, die nicht den im vorangegangenen Kapitel beschriebenen Kriterien für korrekte Videoeinsätze entsprechen, werden daher mit Null bewertet, ebenso alle Sonareinsätze ohne Meeresgrund (> 20 Meter Distanz zum Grund). Für die Bewertung von *Sonareinsätzen auf Grund* werden zwei Situationen unterschieden:

- Sonareinsätze im Zusammenhang mit dem exakten Manövrieren und Ausrichten von Nautilus werden als *Sonar-Folgen* bezeichnet. Zwischen aufeinander folgenden Sonareinsätzen vollzieht der Proband abgesehen von der Steuerung keine anderen Aktionen (Textausgabe „Sonar-Move-Sonar“ oder „Sonar-Video-Sinken-Steigen-Sonar“ etc.). Für solche unmittelbar aufeinander folgenden Sonareinsätze wird die *Effizienz der Verschiebung* beurteilt. Verschiebt der Proband Nautilus nach einem Sonareinsatz *ohne Metallfund* zu wenig (< 0.8 Radius des Sonargerätes), so gewinnt er dadurch keine neue Information, weil er damit zu wenig „neues“ (bisher unge-scanntes) Gebiet prüft. Kleine Verschiebungen (< 0.8 Radius des Sonargerätes) sind nur nach Sonareinsätzen *mit Metallfund* informationseröffnend, da damit der Fund näher geprüft wird.
- Sonareinsätze, zwischen denen auch andere Aktionen ausser der Steuerung vorkommen (lesen, schreiben oder auch Systemmeldungen), werden als *Sonar-Reihen* bezeichnet. Beurteilt wird das Vorliegen von *Wiederholungen*. Neben der Wiederholung durch keine Ortsveränderung gilt auch die alleinige Änderung der Tiefe als Wiederholung, sowie die alleinige Änderung der Richtung (Ausrichtung der Videokamera) nach einem Sonareinsatz ohne Fund oder mit Fund im Scanbereich. Weitere Kriterien zur Berücksichtigung von Wiederholungen bei Sonar- und Videoeinsätzen sind Anhang C zu entnehmen.

3.2.1.3 Kriterien für das Adäquate Handeln

Mit diesem Konstrukt wird erfasst, wie adäquat sich der Proband während der Problembearbeitung verhält. Adäquat meint dabei, dass die einzelnen Handlungsschritte im Sinne der aktuellen Aufgabenstellung sind und zu einer Lösung des

Problems beitragen. Ein adäquates Vorgehen setzt voraus, dass der Proband die wesentlichen Aspekte der Problemstellung versteht und umsetzen kann. Im Allgemeinen bedeutet dies: Der Proband verschafft sich Informationen über das Funktionieren der einzelnen Einheiten des Steuerungscomputers und wendet diese in der dafür vorgesehen Weise korrekt an. Er protokolliert und kommuniziert sein Vorgehen für andere Sucheinheiten transparent und verständlich. Während der Testphase ist es zudem notwendig, dass er die Suchinstrumente (Bewegungsmodi, Sonargerät, Videokamera) nicht nur korrekt einsetzt, sondern die Suche auch effizient gestaltet, da mit steigender Suchzeit die Rettungschancen sinken.

Zum Adäquaten Handeln gehört auch ein weiterer Aspekt: Die einzelnen Handlungsschritte müssen nicht nur der aktuellen Aufgabenstellung entsprechen, sondern auch innerhalb des Handelns des Probanden eine gewisse Logik, das heisst Nachvollziehbarkeit und Konsistenz aufweisen.

Die Bewertung der einzelnen Aktionen erfolgt wie beim Informationseröffnenden Handeln auf einer dreistufigen Skala, die im Unterschied dazu auch eine negative Ausprägung enthält. Die Bedeutung einer Aktion im aktuellen Handlungskontext kann *sinnvoll* (= +1), *wenig sinnvoll* (= 0) oder *nicht sinnvoll* (= -1) sein. Die Einschätzung dieser drei Ausprägungen ist für *alle registrierbaren Aktionen* des Probanden sowie für *einige Systemmeldungen* möglich und bezieht sich wiederum auf die aktuelle Aufgabenstellung und die Vorgeschichte (was der Proband bisher gemacht hat). Für einzelne Aktionen muss auch die unmittelbare Zukunft berücksichtigt werden (was der Proband als nächstes macht).

Sinnvolle Aktionen tragen wesentlich zur Lösung des Problems bei und passen in den aktuellen Handlungskontext. Nicht sinnvolle Aktionen sind Aktionen, die einer effizienten Problemlösung im Wege stehen oder in Bezug auf das bisherige Vorgehen unlogisch oder nicht nachvollziehbar sind. Die mittlere Ausprägung „wenig sinnvoll“ umfasst Aktionen, die weder eindeutig sinnvoll noch eindeutig nicht sinnvoll sind.

Die Bewertung von Aktionen der Handlungskategorie *Lesen* erfolgt nach ähnlichen Überlegungen wie beim Informationseröffnenden Handeln. Sinnvoll (= +1) ist das

erstmalige Lesen von relevanten Informationen. Für eine effiziente Problemlösung wäre es sinnvoll, wenn der Proband beim ersten Lesen die gesamte Information aufnimmt, eine erste Wiederholung gilt deshalb als wenig sinnvoll (= 0), jede weitere Wiederholung als nicht sinnvoll (= -1). Wie beim Informationseröffnenden Handeln gilt eine Seite nur dann als gelesen beziehungsweise wiederholt, wenn die Seite während der minimalen Lesedauer (vgl. Anhang C) geöffnet ist. Ist die Seite weniger lang geöffnet, gilt dies als wenig sinnvoll (= 0).

Für die Informationsseiten in der Datenbank Nauti-Info gilt – im Unterschied zum Informationseröffnenden Handeln – nur das Lesen von *relevanten Informationen* als sinnvoll (= +1), das Lesen von neutralen Informationen wird als wenig sinnvoll (= 0) und das Lesen von unterhaltenden Informationen als nicht sinnvoll (= -1) bewertet. Ein weiterer Unterschied besteht darin, dass bereits das reine Anklicken von Hauptmenü-Seiten unabhängig von der minimalen Lesedauer als sinnvoll gilt (sofern die Seite während mindestens zwei Sekunden geöffnet ist). Dieses rasche Anklicken wird interpretiert als explorative Informationssuche, im Sinne von sich rasch einen Überblick (über die im jeweiligen Menü vorhandenen Funktionen, d.h. Submenüs) verschaffen wollen.

Da das Lesen *eingehender Nachrichten* für die Problemlösung absolut notwendig ist, gilt das Nichtlesen dieser Nachrichten als nicht sinnvoll (= -1), ebenso eine zu grosse Latenzzeit zwischen Ankunft einer Nachricht und Lesezeitpunkt. Auch negativ bewertet wird das Lesen der Trainingsnachrichten 2 und 4, die beide nur verschickt werden, wenn sich der Proband inadäquat verhält (T2: wenn er nach 15 Minuten noch immer kein Protokoll erstellt hat; T4: wenn er sich beim Auftanken nicht am Mutterschiff befindet).

Auch die Bewertung von *Video- und Sonareinsätzen* ist mit wenigen Unterschieden vergleichbar mit den Kriterien für das Informationseröffnende Handeln. Sonareinsätze ohne Grund und nicht korrekte Videoeinsätze sind grundsätzlich nicht sinnvoll (= -1), ebenso Wiederholungen von Sonar- und Videoeinsätzen (vgl. vorhergehendes Kapitel oder Anhang C). Für die Bewertung der *Effizienz der Verschiebung bei Sonar-Folgen* müssen die Distanzangaben etwas präzisiert werden: Wie in Kapitel 3.1.2

erläutert, besteht das optimale Suchvorgehen darin, nach einem *Sonareinsatz ohne Metallfund* Nautilus so zu verschieben, dass der nächste Sonareinsatz möglichst viel neues und möglichst wenig bereits gescanntes Gebiet abdeckt. Als sinnvoll (= +1) gelten deshalb Verschiebungen, die grösser oder gleich 1.2 und kleiner als zwei Radien sind. Verschiebungen von mehr als zwei Radien sind wenig sinnvoll (= 0), weil damit zwischen den beiden Sonareinsätzen Gebiet ungeprüft bleibt (keine Überschneidung der „Kreise“). Nur geringe Verschiebungen (< 0.8 Radius) sind nicht sinnvoll (= -1), weil dadurch keine neue Information aufgenommen wird. Zum optimalen Suchvorgehen gehört auch, nach einem *Sonareinsatz mit Metallfund* diesen näher zu prüfen, das heisst Nautilus so zu manövrieren, dass der Fund in den Sichtbereich der Videokamera rückt. Deshalb sind nach Sonareinsätzen mit Metallfund nur Verschiebungen sinnvoll (= +1), die kleiner als ein Radius sind, Verschiebungen um mehr als einen Radius sind nicht sinnvoll (= -1), weil damit der Metallfund nicht weiter geprüft wird.

Im Unterschied zum Informationseröffnenden Handeln kann beim Adäquaten Handeln auch das *Herumfahren von Nautilus* bewertet werden. Die effizienteste Fortbewegung besteht darin, an der Oberfläche (oder in geringer Tauchtiefe) an den Zielort zu fahren, mit der Steuerautomatik „Grund“ auf drei Meter Distanz zum Meeresboden abzutauchen und unter Beibehalten dieses Abstandes den Grund nach Metallvorkommen abzusuchen. Verzichtet der Proband auf diese automatische Abstandseinhaltung, so muss er entweder ständig die Anzeigen auf dem Monitor überprüfen und Nautilus rechtzeitig sinken oder steigen lassen, oder es kommt zu Kollisionen mit dem Meeresgrund. Unmittelbar vor einer möglichen Kollision erhält der Proband ein Warnsignal und die Weiterfahrt von Nautilus wird blockiert. Erst nach Anpassen der Tauchtiefe kann er Nautilus wieder fortbewegen. Für die Suche nach Metall benötigt die handgesteuerte Fortbewegung daher deutlich mehr Zeit als die automatische Steuerung. Letztere wiederum ist dafür weniger geeignet für das Zurücklegen längerer Wegstrecken, da das andauernde automatische Korrigieren des Abstandes Energie und Zeit braucht.

Bewertet wird vor allem der Einsatz der Steuerautomatik-Funktionen „Grund“ und

„Oberfläche“. So ist beispielsweise das Einstellen der Grundautomatik sinnvoll (= +1), wenn anschliessend Sonar- oder Videoeinsätze (auf Grund) folgen. Das Ausschalten derselben (durch Drücken der Funktion „Sinken“, „Steigen“ oder „Oberfläche“) ist sinnvoll (= +1), wenn anschliessend eine grössere Verschiebung von Nautilus erfolgt (um mehr als einen Radius), jedoch nicht sinnvoll (= -1), wenn anschliessend weitere Sonar- und Videoeinsätze folgen. Weitere Kriterien zum Einsatz der verschiedenen Steuerungsmodi finden sich in Anhang C.

Auch das *Schreiben* von Texteinheiten kann im Hinblick auf das Adäquate Handeln beurteilt werden. Auf der einen Seite lassen sich klar zu bestimmende *formale Bedingungen* definieren, unter denen das Verfassen der Texteinheiten sinnvoll beziehungsweise nicht sinnvoll ist. Nicht sinnvoll (= -1) beispielsweise ist das Verwerfen von Texteinheiten (nicht Speichern von Protokollen, nicht Versenden von Meldungen) oder ineffizientes Schreiben (mehr als zwei Minuten pro Texteinheit). Sinnvoll (= +1) sind alle Protokolle, die auf Verlangen verfasst werden (d.h. aufgrund einer Aufforderung durch die Einsatzleitung).

Auf der anderen Seite muss der *Inhalt der Texteinheiten* qualitativ beurteilt werden. Diese Inhaltsanalyse ist Bestandteil des folgenden Kapitels.

3.2.1.4 Inhalt der verfassten Texteinheiten

Die Beurteilung der Texteinheiten in Bezug auf das Adäquate Handeln erfolgt schrittweise. Zuerst wird die *formale Korrektheit* und *inhaltliche Verständlichkeit* der Angaben geprüft. Ist der Inhalt formal korrekt und verständlich, so wird die *Bedeutung im aktuellen Handlungskontext* eingeschätzt. Nicht verständliche und formal nicht korrekte Texteinheiten werden immer als nicht sinnvoll (= -1) bewertet.

Eine qualitative Inhaltsanalyse ist immer mit einer gewissen Willkür verbunden. Um diese möglichst gering zu halten, wird für die Beurteilung der Texteinheiten auf die mittlere Ausprägung „wenig sinnvoll“ verzichtet. Alle Beurteilungen erfolgen auf einer dichotomen Skala (korrekt versus nicht korrekt, verständlich versus nicht verständlich, sinnvoll versus nicht sinnvoll).

Die *Beurteilung der formalen Korrektheit* betrifft die Fragen, ob die Texteinheit den Vorgaben des Systems entsprechend erstellt wurde, ob alle notwendigen Angaben vorhanden sind und ob diese Angaben auch korrekt sind. In der Einleitung (oder durch Nachlesen in der Bedienungshilfe) erfährt der Proband, dass Protokolle zur Dokumentation des Suchverlaufs und Meldungen zur Kommunikation mit anderen Sucheinheiten verwendet werden sollten. Für die beiden Modi verwendet das System unterschiedliche Eingabemasken: Bei Protokollen muss der Proband die Koordinaten, die aktuelle Tauchtiefe und den Seekartenausschnitt vom Bildschirm ablesen und in den dafür vorgesehenen Feldern eintragen, sowie einen kurzen Text über das Suchergebnis und das weitere Vorgehen verfassen. Meldungen sind ähnlich wie E-Mails aufgebaut und enthalten ein Feld für ein Stichwort (den Betreff), ein Feld für den Status (z.B. dringend) sowie das eigentliche Textfeld. Über ein Submenü kann nach Beendigung des Schreibens der Adressat ausgewählt werden.

Der *Modus „Protokoll“* gilt als korrekt, wenn das Protokoll alle Angaben über die Position von Nautilus sowie einen stichwortartigen Text über den aktuellen Stand der Suche enthält. Enthält ein Protokoll Fragen oder Mitteilungen an andere Sucheinheiten, so gilt der Modus als falsch. Der *Modus „Meldung“* gilt als korrekt, wenn die Meldung eine Mitteilung an andere Sucheinheiten enthält und der Adressat mit dem Inhalt übereinstimmt (die Auftankmeldung z.B. muss an die Technik und nicht an die Einsatzleitung verschickt werden). Enthält eine Meldung keinen Text oder nur Positionsangaben ohne weitere Informationen, so gilt der Modus als falsch. Neben dem Modus müssen auch *alle Angaben über die Position von Nautilus korrekt* eingetragen sein, das heisst Nautilus muss sich zum Zeitpunkt des Verfassens der Texteinheit wirklich an der genannten Stelle befinden (oder falls sich der Inhalt auf vorausgegangene Sonar- oder Videoeinsätze oder andere Aktionen bezieht, müssen die Angaben mit dieser Position übereinstimmen).

Ist eine Texteinheit formal korrekt erstellt, so wird als nächstes die *inhaltliche Verständlichkeit* geprüft. Verständliche Texteinheiten enthalten für einen Dritten nachvollziehbare Informationen und lassen den Gedankengang des Probanden klar erkennen. Unverständliche Texteinheiten sind meist lückenhaft. Es bleibt unklar, welche Überlegungen der Proband festhalten oder übermitteln will.

Bei inhaltlich verständlichen und formal korrekt erstellten Texteinheiten wird die *Bedeutung im aktuellen Handlungskontext* eingeschätzt. Sinnvolle Texteinheiten (= +1) sind konstruktiv, tragen zur Problemlösung bei und enthalten für andere Sucheinheiten wichtige Informationen, beispielsweise Angaben über das Suchergebnis oder das weitere Vorgehen. Der Inhalt muss zum momentanen Vorgehen des Probanden und zum aktuellen Stand der Suche passen. Beispielsweise ist die Information „Habe im Umkreis von 160 Metern nichts festgestellt“ wohl für andere Sucheinheiten relevant, wenn davor jedoch keine Sonar- und Videoeinsätze gemacht wurden, so ist die Aussage nicht zutreffend und damit für andere irreführend. Sind die Aussagen im aktuellen Handlungskontext unlogisch, nicht passend, oder enthält der Text nur unterhaltende, für die Problemlösung irrelevante oder gar falsche Informationen, so gilt die Texteinheit als nicht sinnvoll (= -1). Ebenfalls als nicht sinnvoll zählen Texteinheiten mit dem gleichen Inhalt wie die vorangegangene, im gleichen Modus erstellte Texteinheit.

Texteinheiten, deren Bedeutung im aktuellen Handlungskontext über das normal Erwartete hinausreicht, werden mit einem zusätzlichen Punkt bewertet. Als *kreativ* oder *aussergewöhnlich* gilt beispielsweise die (korrekte) Begründung für eine Suche in Gebieten mit einer Tiefe von mehr als 50 Metern oder die Anforderung weiterer Helfer. Der Proband denkt mit und schlägt mögliche (neue) Wege vor, wie das Problem schneller gelöst werden kann. Diese zusätzliche Bewertung von kreativen Ideen soll dem Sachverhalt Rechnung tragen, dass Probanden, die gut und konstruktiv kommunizieren, für diesen (zeitlichen) Aufwand relativ wenig „belohnt“ werden. Im Mittel schreiben die Probanden 20 Texteinheiten, die Gesamtzahl an registrierten Aktionen liegt jedoch im Mittel bei knapp 300.

Zur Veranschaulichung dieser Kriterien sind in Tabelle 3 ein paar Beispiele wiedergegeben.

Tabelle 3: Beispiele für die Bewertung von Textinhalten.

Wortgetreue Wiedergabe der Texteinheiten	Erklärung
<i>Beispiele für nicht sinnvolle Texteinheiten</i>	
Message Nr. 9 Jetset wichtig Bestaetigen nach erhaltener Meldung den Abbruch unserer Suche. Gehen auf Forschungsreise nach T9 auf Suche nach Erzhaltigem Gestein. Melden uns wieder. -> an alle	Der Proband hat soeben die Meldung 4 gelesen, die darauf hinweist, dass das Flugzeug in der Gegend R/9 abgestürzt sein könnte.
Protokoll Nr. 4 289/43 R/8 118 Habe nicht definierbares Metalteil entdeckt. Bitte Taucher an angegebene Koordina gehe in Planquadrat 9/R	Das Protokoll enthält eine Mitteilung an die Tauchergruppe. Im Anschluss folgt keine entsprechende Meldung an dieselbe.
Message Nr. 10 Abendessen dringend termin planen 20 uhr mit herrx -> an Terminplanung	Für die Problemlösung absolut irrelevant (der Proband hat soeben die Meldung 3 gelesen).
<i>Beispiele für sinnvolle Texteinheiten</i>	
Message Nr. 5 jet vermisst Warten mit der Tauchequippe. Habe Photos gemacht, sieht nicht aus wie ein Flugzeug. Suche weiter und Informiere -> an Tauchergruppe	Der Proband antwortet verständlich auf die soeben gelesene Nachricht der Taucher. Die davor erstellten Sonar- und Videoaufnahmen waren korrekt, ergaben jedoch keine positiven Befunde.
Message Nr. 4 Energie wichtig Nautilus hat noch 40% Energie und ist auf dem Weg zu R9. Erbitte Koordination mit anderen Schiffen -> an Leitung	Der Proband realisiert den schwindenden Energiestand (nach Lesen von Meldung 4), kommuniziert sein weiteres Vorgehen und verlangt die Koordination mit anderen Sucheinheiten. (<i>kreativ</i>)
Message Nr. 3 Wrackteile in Abklärung Taucher bereithalten für schnellen Einsatz Sind weitere Erkenntnisse vorhanden? -> an alle	Der Proband hat soeben auf dem Videobild Gepäckstücke gesehen und kommt von alleine auf die Idee, die Taucher anzufordern (die Systemmeldung „Anfrage der Taucher“ erhält er erst später). Zudem fragt er andere Sucheinheiten nach neuen Erkenntnissen. (<i>kreativ</i>)

3.2.2 Prozessdarstellung und -beschreibung

Die nach der Mikroverhaltensbeschreibung vorliegenden Werte werden über die Zeit aufsummiert und die kumulierten Werte als Zeitreihen dargestellt. Für die graphische Darstellung wird ein von Wehrle⁵ entwickeltes Programm verwendet. Die Zeitreihen werden jeweils für die *gesamte Bearbeitungszeit* (Trainings- und Testphase) dargestellt und bewertet, weil auch die szenariospezifische Auswertung für die Beurteilung der Problemlöseleistung beide Spielphasen berücksichtigt. Der Verlauf der Testphase wird zudem auch separat, das heisst unabhängig vom Verlauf während der Trainingsphase betrachtet (durch Rekalibrierung, d.h. Nullsetzung der Zeitreihe bei Trainingsende).

Bei der *qualitativen Beschreibung* der Zeitreihen interessiert, welche Charakteristika die Verläufe der beiden Konstrukte aufweisen und worin sie sich unterscheiden. Aufgrund bisheriger Studien oder Erfahrungen mit ATLANTIS können keine Hypothesen formuliert werden, welche Kurvenverläufe am ehesten zu erwarten sind. Unklar ist auch, welche Kombinationen der Verläufe für das Informationseröffnende und Adäquate Handeln vorkommen und wie häufig. An dieser Stelle soll deshalb nur kurz überlegt werden, welche Verläufe theoretisch, das heisst aufgrund der gewählten Codierung der Mikroverhaltensdaten möglich sind.

Für das *Informationseröffnende Handeln* bedeutet die Codierung, dass die (potentielle) Informationsmenge des Probanden nur zunehmen oder stagnieren, jedoch nicht abnehmen kann. Die Kurven können damit nur mehr oder weniger steil sowie mehr oder weniger konstant ansteigen. Flache Abschnitte in der Kurve („Plateaus“) bedeuten, dass der Proband in dieser Zeit keine problemrelevanten Informationen aufgenommen hat. Beim *Adäquaten Handeln* sind aufgrund der negativen Codierung von „nicht sinnvoll“ auch weitere Verläufe möglich: negative (abfallende) Verläufe oder wechselnde Verläufe, wie beispielsweise „mehr oder weniger konstant um Null schwankend“ oder „zuerst konstant ansteigend, dann abfallend“ oder umgekehrt „erst abfallend, dann ansteigend“ usw.

⁵ Dr. Thomas Wehrle ist wissenschaftlicher Mitarbeiter am Psychologischen Institut der Universität Zürich.

Die *quantitative Beschreibung* der Zeitreihen betrifft die Definition eines Prozessmasses, das den Kurvenverlauf formal beschreibt. Es geht dabei nicht um die mathematisch-statistische Analyse der Zeitreihen, sondern um eine einfach quantifizierbare, an inhaltlichen Überlegungen orientierte Beschreibung des Kurvenverlaufs. Durch die *Ausprägungsgrade der Zeitreihe* (d.h. Höhe der Kurve) zu *bestimmten Zeitpunkten* soll der Kurvenverlauf abgebildet werden. Mit Zeitpunkten sind jedoch nicht absolute Zeitpunkte gemeint, da die Probanden zum gleichen Zeitpunkt in unterschiedlichen Spielphasen sein können. In der gleichen Spielphase und damit auf dem gleichen „Informationsstand“ befinden sich die Probanden bei den vom System vorgegebenen Ereignissen. Tabelle 4 gibt eine Übersicht über alle möglichen *Systemereignisse* in ATLANTIS und die in dieser Arbeit verwendeten Abkürzungen für diese Ereignisse.

Von allen möglichen Systemereignissen werden die gewählt, die grundsätzlich bei allen Probanden vorkommen und die nicht unmittelbar aufeinander folgen: die Systemmeldungen *T3*, *T*, *M2*, *M3*, *M4*, sowie der *Beginn der Testphase* (*X*) und das *Spielende*. Die Systemmeldung *M1* wird nicht berücksichtigt, da diese nur wenige Sekunden nach Beginn der Testphase erscheint, und daher die Ausprägungsgrade der beiden Konstrukte zu den Zeitpunkten *X* und *M1* nur minimal von einander abweichen. Das Gleiche gilt für den Zeitpunkt *ET* (Trainingsende), der sich nur maximal drei Minuten vor *X* ereignet.

Die Ausprägungsgrade der jeweiligen Zeitreihe (*Info*, *Ad*) bei den Systemereignissen *T3*, *X*, *M2*, *M3*, *M4* und Ende bilden das *Prozessmass*. Das Prozessmass ist als *n*-Tupel oder Vektor zu verstehen. Wenn im Folgenden von den Komponenten des Prozessmasses die Rede ist, wird dafür der Begriff *Ereignispunkte* verwendet.

Für eine korrekte Beschreibung und Interpretation des Kurvenverlaufs muss natürlich auch die *Spieldauer* berücksichtigt werden. Wie bei der Beschreibung von ATLANTIS bereits erwähnt, variiert diese je nach Spielverhalten beträchtlich (vgl. Kapitel 3.1.2). In Ergänzung zum Prozessmass für das Adäquate und Informationseröffnende Handeln werden deshalb auch die *Gesamtspieldauer*, die *Trainings-* und die *Testzeit* verwendet, um den Kurvenverlauf quantitativ zu beschreiben.

Beide zusammen, Prozessmass und Angaben zur Spieldauer, spielen für die Definition der Prozessprofile eine Rolle, wie das folgende Kapitel zeigt.

Tabelle 4: Übersicht über die chronologisch geordneten Systemereignisse in ATLANTIS und die in dieser Arbeit verwendeten Abkürzungen.

Abkürzung	Beschreibung der Systemereignisse
T1	Trainingsmeldung „Protokollerinnerung“ <i>Bedingung:</i> Nach 10 Minuten noch kein Protokoll erstellt
T2	Trainingsmeldung „Protokollaufforderung“ <i>Bedingung:</i> Nach 15 Minuten noch kein Protokoll erstellt
T3	Trainingsmeldung „Auftanken!“ <i>Bedingung:</i> Erstes Protokoll erstellt
T4	Trainingsmeldung „Positionskorrektur“ <i>Bedingung:</i> Auftankmeldung verschickt, aber Nautilus ist falsch platziert
ET	Trainingsende (ohne Auftankvorgang)
X	Ende Auftankvorgang, Beginn Testphase
M1	Erste Notfallmeldung
G	Sichtung des Gepäckstückes <i>Bedingung:</i> 6-9 Minuten nach Beginn der Testphase in S/7 oder R/9 Meeresbilder vom Grund erstellt
T	Anfrage der Taucher
A	Antwort der Taucher <i>Bedingung:</i> Anfrage der Taucher mit Meldung beantwortet
M2	Zweite Notfallmeldung
M3	Dritte Notfallmeldung
M4	Vierte Notfallmeldung (Ort sensitiv)
F	Sichtung von Wrackteilen des Flugzeuges <i>Bedingung:</i> 5-11 Minuten nach M4 Meeresbilder vom Grund erstellt
M5	Fünfte Notfallmeldung (Ende 1) <i>Bedingung:</i> Kein F, oder nach F kein Protokoll bzw. keine Meldung erstellt
M6	Sechste Notfallmeldung (Ende 2-1) <i>Bedingung:</i> Nach F Protokoll bzw. Meldung erstellt
M7	Siebte Notfallmeldung (Ende 2-2) <i>Bedingung:</i> Nach Erhalt von M6
Ende	Spielende

3.2.3 Beurteilung der Ähnlichkeit der Zeitreihen

Die Definition von Prozessprofilen erfolgt durch die Beurteilung der Ähnlichkeit der Zeitreihen. Ziel ist, für jedes Konstrukt möglichst wenige, klar voneinander abgrenzbare typische Verläufe (Prozessprofile) zu finden. Probanden aus dem gleichen Profil sollten möglichst ähnliche, Probanden aus verschiedenen Profilen möglichst unterschiedliche Verläufe aufweisen. Für die Beurteilung der Ähnlichkeit der Zeitreihen werden verschiedene Klassifizierungszugänge verwendet, die auf unterschiedlichen inhaltlichen und formalen Kriterien beruhen und die in unterschiedlicher Weise die Ereignispunkte (das Prozessmass) und die Spieldauer berücksichtigen.

In einem ersten Schritt wird die Ähnlichkeit der Zeitreihen nach *inhaltlichen Überlegungen* beurteilt. Massgebend ist dabei nicht die absolute Ähnlichkeit der Zeitreihen (z.B. möglichst deckungsgleich), sondern die Ähnlichkeit der Verläufe in Bezug auf den *Spielverlauf*, das heisst in Bezug auf die Systemereignisse. Die absolute Spielzeit spielt eher eine untergeordnete Rolle. Nach der Bestimmung möglicher Prozessprofile werden Kriterien definiert, die die Zugehörigkeit zu einem Profil eindeutig festlegen. Die Definition dieser Zuordnungsregeln oder *Klassifikationsregeln*⁶ mag zu Recht willkürlich erscheinen, zudem ist sie in hohem Masse stichprobenabhängig: Es gibt immer Probanden, die die Klassengrenzen knapp verfehlen. Aus diesem Grunde werden die Klassifikationsregeln auch an einem Datensatz angewendet und überprüft, der nicht für die Klassifizierung verwendet wurde.

Ausserdem wird die Willkür beziehungsweise die Plausibilität dieser inhaltlich orientierten Klassifizierung durch die Anwendung *formaler Klassifizierungsmethoden* überprüft. Dabei werden zwei grundsätzlich verschiedene Ansätze verfolgt: Ansatz A basiert, analog zur qualitativen Beurteilung, vorwiegend auf den Systemereignissen, während Ansatz B die absolute Spielzeit und damit die absolute Ähnlichkeit der Zeitreihen berücksichtigt.

⁶ An dieser Stelle wird bewusst der Begriff „Klassifikation“ statt „Klassifizierung“ verwendet, weil damit sowohl die Klassenbildung wie auch die Klassenzuordnung bezeichnet wird (Kallus & Janke, 1995).

Beim *Ansatz A* werden *Clusteranalysen über das Prozessmass* gerechnet. Es werden für jedes Konstrukt die Zeitreihen zu einem Cluster zusammengefasst, die zum Zeitpunkt der ausgewählten Systemereignisse ähnliche Ausprägungsgrade haben. Die Spieldauer wird mit in die Clusteranalyse einbezogen, falls diese auch bei der inhaltlich orientierten Klassifizierung eine Rolle spielt.

Beim *Ansatz B* wird für jede Zeitreihe der „*Abstand*“ zu allen anderen Zeitreihen berechnet. Ein grosser Abstand steht für geringe Ähnlichkeit zwischen zwei Zeitreihen und umgekehrt. Aus der Berechnung der Abstandsmasse resultiert eine Ähnlichkeitsmatrix, anhand derer mittels Multidimensionaler Skalierung und Clusteranalysen die Gruppenbildung erfolgt.

Als Mass für den Abstand zwischen zwei Zeitreihen wird die *Euklidische Distanz* gewählt. Diese ist definiert als die Wurzel aus den quadrierten und aufsummierten Differenzwerten, wobei dieser Wert in Abhängigkeit von der Anzahl Messzeitpunkte (zur Berechnung der Differenzwerte) unterschiedlich exakt sein kann. Erschwerend für die Berechnung⁷ der Euklidischen Distanzen ist, dass die Zeitreihen unterschiedlich lang sind. Drei verschiedene Berechnungsvarianten sollen diesem Umstand Rechnung tragen:

1. Variante „*Fixed number of measures*“: Es wird das gemeinsame Intervall zwischen zwei Zeitreihen bestimmt und über dieses Intervall werden n Messpunkte gelegt. Sehr kurze Zeitreihen führen zu Verzerrungen, da der nicht im gemeinsamen Intervall liegende Bereich der längeren Zeitreihe ignoriert wird.
2. Variante „*Fixed measure interval / Fixed Step*“: Die Grösse des Intervalls wird bestimmt durch die längere der beiden Zeitreihen, wobei die kürzere Zeitreihe mit konstant bleibendem Endwert verlängert wird. Über dieses Intervall wird in konstanten Abständen gemessen.
3. Variante „*Adjust X Scale*“: Beide Zeitreihen werden auf ein Normintervall abgebildet und über dieses Intervall n Messzeitpunkte gelegt.

Während die ersten beiden Varianten auf unterschiedliche Weise die Varianz in der Spielzeit berücksichtigen, ist Variante 3 von der Spieldauer unabhängig. Verglichen wird nur die Verlaufsform zweier Zeitreihen, unabhängig davon, wie lange die beiden Probanden gespielt haben.

⁷ Die Berechnung der Euklidischen Distanzen erfolgt mit dem auf Seite 81 erwähnten Programm von Wehrle.

Die Euklidische Distanzmatrix wird nun einerseits als Ähnlichkeitsmatrix betrachtet, die Ausgangspunkt der *Multidimensionalen Skalierung (MDS)* ist. Ziel der MDS ist hier allerdings nicht, die Ähnlichkeiten zwischen den Zeitreihen durch wenige Dimensionen zu beschreiben. Es wird lediglich versucht, die Ähnlichkeiten (d.h. die berechneten Euklidischen Distanzen) zwischen den Zeitreihen durch ihre Distanzen in einem zweidimensionalen Darstellungsraum bestmöglich darzustellen. Der ermittelten Konfiguration soll entnommen werden, welche Zeitreihen „nahe beieinander liegen“ und aufgrund dieser Ähnlichkeiten am ehesten zu Prozessprofilen zusammengefasst werden können. Die inhaltliche Bedeutungszuweisung der beiden Ordnungsdimensionen steht nicht im Vordergrund.

Andererseits werden die berechneten Euklidischen Distanzen zwischen den Zeitreihen als Ortsvektoren im n-dimensionalen Raum verstanden (wobei n die Anzahl Zeitreihen, d.h. Probanden bezeichnet) und über diese *Clusteranalysen* gerechnet. Im Unterschied zur MDS kann der Clusteranalyse eindeutig eine Gruppenbildung entnommen werden.

3.3 Beschreibung des Expertenurteils

Es wird ein Experte gewählt, der mit der Durchführung und Auswertung von ATLANTIS seit Jahren bestens vertraut ist⁸. Der Experte erhält die Aufgabe, die Probandenleistung in Bezug auf die durch das Auswertungsmodell definierten Konstrukte einzuschätzen. Dabei wird ihm lediglich folgende knappe Beschreibung der beiden Konstrukte vorgelegt:

- *Informationseröffnendes Handeln*: Dieses Konstrukt erfasst, wie gut sich der Proband während der Bearbeitung immer wieder neue, bezüglich der aktuellen Problemstellung relevante Informationen verschafft, das heisst wie sich der Proband hinsichtlich „Informationszuwachs“ verhält.
- *Adäquates Handeln*: Dieses Konstrukt erfasst, wie sinnvoll die einzelnen Handlungsschritte im aktuellen Handlungskontext des Probanden sind, das heisst inwiefern sich der Proband der Problemstellung entsprechend und konsistent verhält.

⁸ Aus Datenschutzgründen bleibt der Experte in dieser Arbeit anonym.

Die Prozessprofile der beiden Konstrukte bilden die (ordinal gestuften) Urteilskategorien, wobei dem Experten nur deren Anzahl und grobe Bedeutung vorgegeben wird (z.B. 1 = sehr gute, 2 = gute, 3 = schlechte Informationsaufnahme). Zur Beurteilung der Problemlöseleistung werden dem Experten nur die Protokollbogen und die Graphiken über den Suchweg und die Suchstrategie zur Verfügung gestellt, nicht jedoch die Verhaltensprotokolle. Dies entspricht der Datenbasis, die im Anwendungsfeld der Personalauswahl und -entwicklung auch den Diagnostikern zur Leistungsbeurteilung zur Verfügung steht (vgl. Kapitel 3.1.3.3).

Der Experte beurteilt das Problemlöseverhalten gemäss eigenen Angaben aufgrund folgender Überlegungen:

- *Schaffen einer Handlungsbasis*: Hat der Proband das Training erfüllt? Begründungen? Charakteristik dieser Phase?
- *Generelle Problemerkennung und -bearbeitung*: Wie ist die Gesamtbewegung von Nautilus? Ist die Bewegung von Nautilus effizient (Grund, Oberfläche,...)? Ist die Informationsaufnahme gewährleistet (Ausmass an Aktivität, Menge, Gebiete,...)?
- *Konkrete Schritte*: Bewertung der 13 Leistungskennwerte (Vorgaben erreicht? In welchen Kennwerten? Anteil versus Anzahl?) und Bewertung der Texte insgesamt.
- *Kommunizierte Problemsicht*: Sind die Texte a) verständlich, b) sprachlich korrekt, c) logisch, d) kreativ? Was lässt sich aufgrund der Texte über das generelle Modell des Probanden aussagen? Was lässt sich über die verbale Gesamtcharakteristik aussagen?

Zu diesen Punkten trägt der Experte in einem Excel-File für jeden Probanden stichwortartig seine Beobachtungen und Einschätzungen ein und integriert anschliessend die verschiedenen, daraus resultierenden Informationen über die Güte der Problembearbeitung zu einem Gesamturteil (bezüglich den Info- und Ad-Urteilskategorien). Wie genau diese Datenintegration erfolgt, ob dabei implizite Regeln oder Gewichtungen einzelner Aspekte angewendet werden, ist nicht bekannt.

3.4 Vorgehen und Stichprobe

Wie bereits bei den Fragestellungen erwähnt, wird an einer *ersten Stichprobe* das Beurteilungsschema entwickelt, das heisst die Interpretationsregeln der Mikroverhaltensbeschreibung sowie die Klassifikationsregeln zur Definition der Prozess-

profile. An einer *zweiten, von dieser Entwicklung unabhängigen Stichprobe* werden diese Regeln dann angewendet. Das bedeutet: Die zweite Stichprobe wird erst *nach* Abschluss der Untersuchungen an der ersten Stichprobe zusammengestellt und nach den Regeln des neuen Auswertungsmodells bewertet.

Das *Expertenurteil* erfolgt nach Abschluss dieser Bewertungen, um sicher zu stellen, dass die Ergebnisse des Expertenurteils nicht die eigene Arbeit beeinflussen. Zuerst erhält der Experte die Daten der ersten Stichprobe. Nach Rückgabe der bewerteten Daten werden ihm die Daten der zweiten Stichprobe zugestellt. Die Überprüfung des Auswertungsmodells (d.h. der Vergleich mit der szenariospezifischen Auswertung und dem Expertenurteil) erfolgt nach Erhalt des vollständigen Expertenurteils.

Für die Entwicklung des neuen Auswertungsmodells kann auf einen grossen Datensatz zurückgegriffen werden: Im Rahmen der Personalauswahl des bereits erwähnten Dienstleistungsunternehmens wurden zwischen 1999 und 2005 über 200 Probanden mit ALTANTIS getestet. Insgesamt stehen 230 verwertbare und interpretierbare Datensätze zur Verfügung, die im Folgenden als *Gesamtstichprobe* bezeichnet werden. Bei 198 beziehungsweise 199 Personen liegen auch Angaben zum Geschlecht, zum Alter und zur Ausbildung vor. Demnach waren die Teilnehmer vorwiegend männlich (83% zu 17%) und zwischen 24 und 59 Jahren alt ($M = 39$, $SD = 8$). 19% haben eine Lehre absolviert, 5% die Mittelschule, 27% verfügen über einen Hochschulabschluss, 32% über eine Fachausbildung und 17% haben das (Abend-) Technikum besucht.

Aus dieser Gesamtstichprobe werden die beiden Stichproben zur Entwicklung und Überprüfung des neuen Auswertungsmodells generiert. *Teilstichprobe 1* wird nach inhaltlichen Überlegungen so zusammengesetzt, dass möglichst das ganze Spektrum an registrierten Verhaltensweisen während der Szenarienbearbeitung erfasst wird. Dazu werden die 230 Probanden und Probandinnen der Gesamtstichprobe anhand der *graphischen Darstellung des Suchwegs und der Suchstrategie* (vgl. Kapitel 3.1.3.1) nach folgenden Kriterien beurteilt:

- Trainingsphase: Nautilus in T9
- Anzahl Sonar- und Videoeinsätze (auf Grund, nicht auf Grund)
- Suche im linken (S/7) und rechten (R/9) Seekartenausschnitt

- Suche nur in S/7 oder nur in R/9
- Beginn der Suche in R/9
- Suche in keinem der beiden Gebiete mit Meerestiefe < 50m
- Anzahl Wechsel des Suchgebiets (0, 1, 2, >2)
- Suchaktivität (keine, wenig, mittel, viel)

Diese Beurteilungen erfolgen nicht systematisch. Es soll lediglich versucht werden, möglichst viele unterschiedliche Problembearbeitungen und Suchstrategien ausfindig zu machen und innerhalb solcher Gruppen wiederum möglichst verschiedene Probanden auszuwählen (z.B. Probanden, die mit der Suche in R/9 beginnen, wovon einer ohne anschließenden Wechsel des Suchgebiets, einer mit Wechsel etc.). Nach diesem Vorgehen wird eine Stichprobe von 35 Personen zusammengestellt.

Die Zusammenstellung der *Teilstichprobe 2* erfolgt nach dem Zufallsprinzip, wobei pro Erhebungsjahr zufällig je 3 Personen ausgewählt wurden. Dies ergibt eine Stichprobe von 21 Personen.

3.5 Auswertungsverfahren

Um zu beurteilen, ob der neue Ansatz die individuellen Unterschiede im Problemlöseverhalten aufzeigen kann (*Fragestellung 1*), werden die beobachteten Zeitreihen qualitativ und quantitativ (anhand deskriptiver Statistiken zu den Prozessmassen) beschrieben. Die Definition der Prozessprofile (*Fragestellungen 2 und 3*) erfolgt entsprechend den Erläuterungen über die Beurteilung der Ähnlichkeit der Zeitreihen (vgl. Kapitel 3.2.3). Auf Auswertungsproblematiken im Zusammenhang mit den *formalen Klassifizierungsmethoden* (Clusteranalyse, Multidimensionale Skalierung) wird im Ergebnisteil eingegangen.

Die *vierte Fragestellung* überprüft, ob die Problemlöseleistung durch das neue Auswertungsmodell ähnlich gut eingeschätzt wird wie durch den szenariospezifischen Auswertungsansatz. Die Operationalisierung der Problemlöseleistung erfolgt beim szenariospezifischen Ansatz durch die 13 Leistungskennwerte (vgl. Tabelle 2, Seite 63), beim neuen Auswertungsmodell durch die Prozessmasse und Prozessprofile für

das Informationseröffnende und Adäquate Handeln. Zur Beurteilung des Zusammenhangs zwischen den beiden Ansätzen werden durchgehend *nicht parametrische Auswertungsverfahren* angewendet, da alle 13 Leistungskennwerte in der Gesamtstichprobe nicht normalverteilt sind.

Die Prüfung der *Hypothese 4.1*, die einen positiven Zusammenhang zwischen den Prozessmassen und den Leistungskennwerten postuliert, erfolgt mittels *bivariaten Korrelationen* bei einseitiger Hypothesenprüfung. Da die Leistungskennwerte aus einer Aufsummierung bestimmter Verhaltensaspekte über die gesamte Bearbeitungszeit bestehen und damit globale Masse bilden, wird bei den beiden Prozessmassen nur die Komponente *Endwert* berücksichtigt.

Für die Prozessprofile wird erwartet, dass sich diese mit monoton fallendem Trend in Bezug auf die Leistungskennwerte unterscheiden (*Hypothese 4.2*). Anhand des *Trendtests nach Jonckheere-Terpstra* soll diese Hypothese geprüft werden. Der Trendtest nach Jonckheere-Terpstra ermöglicht, auch bei nicht normalverteilten Daten und mehreren Stichproben, Trendhypothesen zu formulieren. Geprüft wird die Trendhypothese, dass die Medianwerte der Prozessprofile einer schwach monotonen Rangordnung folgen (wobei Profil $1 > 2 > 3$ etc.).

Mit den letzten beiden *Fragestellungen 5 und 6* soll beantwortet werden, ob die Einschätzungen der Problemlöseleistung durch das neue Auswertungsmodell mit den Einschätzungen eines Experten übereinstimmen (wobei der Experte die Probandenleistung auf ordinal gestuften Urteilskategorien einschätzt, die der Anzahl und Rangordnung der Prozessprofile entsprechen). Dazu wird einerseits der Zusammenhang zwischen den Prozessprofilen und dem Expertenurteil mittels *bivariater Korrelation* beurteilt (einseitige Hypothesenprüfung), andererseits wird die Urteilskonkordanz mit dem „*weighted-kappa*“-Koeffizienten (κ_w) von Cohen (1968) beurteilt. Der „*weighted-kappa*“-Koeffizient bietet im Unterschied zum normalen Kappa-Koeffizienten die Möglichkeit, bei nicht übereinstimmenden (d.h. diskordanten) Urteilen den Grad beziehungsweise das Ausmass der Nichtübereinstimmung mit zu berücksichtigen (bzw. analog für konkordante Urteile den Grad der Übereinstimmung). Für die vorliegende Untersuchung erscheint eine solche Gewichtung der

Diskordanz sinnvoll, da es weniger gravierend ist, wenn ein Proband beispielsweise (bei vier Prozessprofilen) Rang 4 statt Rang 3 zugeordnet wird, als wenn er Rang 4 statt Rang 1 zugeordnet wird. Diskordanzen in benachbarten Urteilskategorien sind damit weniger stark zu gewichten als Diskordanzen in auseinander liegenden Kategorien. Zudem sind Diskordanzen im unteren Leistungsbereich eher in Kauf zu nehmen als im oberen Leistungsbereich, da ATLANTIS als Selektionsinstrument angewendet wird, mit dem Ziel, die für Kaderpositionen geeignetsten Personen zu identifizieren.

Zur Berechnung des gewichteten Kappa-Koeffizienten werden die Urteile der beiden Beurteiler (hier die Beurteilung durch das neue Auswertungsmodell und den Experten) in einer Kreuztabelle (Kontingenztafel) eingetragen. Die Summe der Häufigkeiten in der Hauptdiagonale gibt die Anzahl der konkordanten, die Summe aller Häufigkeiten ausserhalb der Diagonale die Anzahl der diskordanten Urteile wieder. Im Unterschied zum normalen Kappa-Koeffizienten wird nun jede Zelle der $k \times k$ -Tafel mit einem *Diskrepanzgewicht* v_{ij} versehen (wobei $v_{ij} = v_{ji}$, da die beiden Urteile als äquivalent betrachtet werden), und diese Gewichte mit den beobachteten (f_{ji}) sowie den bei stochastischer Unabhängigkeit der beiden Urteile theoretisch erwarteten (e_{ij}) Häufigkeiten multipliziert (für die Berechnung von κ_w vgl. Bortz & Lienert, 2003, S. 297). Grundsätzlich können Diskrepanzgewichte beliebige Zahlen sein, zur besseren Vergleichbarkeit empfehlen Bortz und Lienert (ebd.) jedoch, das Intervall von 0 bis 1 zu wählen. Konkordante Urteile erhalten das Gewicht 0, die diskrepananten Urteile werden umso höher gewichtet, je grösser die Distanz zwischen den Urteilskategorien ist, wobei die maximal mögliche Distanz das Gewicht 1 erhält. Für diese Arbeit sollen die Diskrepanzgewichte so gewählt werden, dass Abweichungen von der besten zur schlechtesten Urteilskategorie vier Mal so stark gewichtet werden wie Abweichungen von der schlechtesten zur zweitschlechtesten und doppelt so stark wie Abweichungen von der zweitbesten zur besten Kategorie. Da an dieser Stelle noch unklar ist, wie viele und welche Prozessprofile (und damit Urteilskategorien für das Expertenurteil) sich überhaupt definieren lassen, sollen die Diskrepanzgewichte für eine drei bis fünf Gruppenlösung definiert werden:

- Bei drei Profilen, wobei 1 die beste, 3 die schlechteste Kategorie bezeichnet (im Sinne von „gut“, „mittel“, „schlecht“):
 - Diskordanz 1-3 $\rightarrow 1$
 - Diskordanz 1-2 $\rightarrow 0.5$
 - Diskordanz 2-3 $\rightarrow 0.25$
- Bei vier Profilen, wobei 1 die beste, 4 die schlechteste Kategorie bezeichnet (im Sinne von „sehr gut“, „gut“, „mittel“, „schlecht“):
 - Diskordanz 1-4 $\rightarrow 1$
 - Diskordanz 1-3 und 2-4 $\rightarrow 0.75$
 - Diskordanz 1-2 und 2-3 $\rightarrow 0.5$
 - Diskordanz 3-4 $\rightarrow 0.25$
- Bei fünf Profilen, wobei 1 die beste, 5 die schlechteste Kategorie bezeichnet (im Sinne von „sehr gut“, „gut“, „mittel“, „schlecht“, „sehr schlecht“):
 - Diskordanz 1-5, 1-4 und 2-5 $\rightarrow 1$
 - Diskordanz 1-3, 2-4 und 3-5 $\rightarrow 0.75$
 - Diskordanz 1-2 und 2-3 $\rightarrow 0.5$
 - Diskordanz 3-4 und 4-5 $\rightarrow 0.25$

Neben der Urteilskonkordanz in Bezug auf die Prozessprofile der beiden Konstrukte soll auch die Übereinstimmung in Bezug auf die daraus resultierenden Problemlösetypen (im Beurteilungsschema) betrachtet werden. Hier interessiert vor allem die Übereinstimmung im oberen Leistungsbereich, während für den unteren, weniger gut differenzierenden Bereich mehr Diskrepanzen in Kauf genommen werden. Es werden nur absolute und relative Häufigkeiten für die gesamte Tafel beziehungsweise einzelne Felder berechnet, da die Probandenleistung nicht in Bezug auf die Problemlösetypen eingeschätzt werden, sondern in Bezug auf die Prozessprofile beziehungsweise Urteils Kategorien der beiden einzelnen Konstrukte.

Bevor nun in den folgenden Abschnitten die Ergebnisse dieser Berechnungen präsentiert werden, noch ein paar Anmerkungen zu den verwendeten Auswertungsprogrammen: Die Codierung der Verhaltensprotokolle (Mikroverhaltensbeschreibung) erfolgte in Excel. Für die graphische Darstellung der Zeitreihen und die Berechnung der Euklidischen Distanzen wurde das auf Seite 81 erwähnte Programm von Wehrle verwendet. Die statistischen Auswertungen wurden mit dem Programm SPSS Version 14 durchgeführt.

4 Ergebnisse und erste Interpretation

Die Ergebnisse dieser Arbeit werden entsprechend der Reihenfolge der Fragestellungen (vgl. Kapitel 2.4) dargestellt. Da die Fragestellungen aufeinander aufbauen, werden einzelne Ergebnisse, wo es für das Verständnis hilfreich ist, bereits in diesem Teil interpretiert. Eine abschliessende und zusammenfassende Diskussion folgt in Kapitel 5.

Bevor die Ergebnisse zu den einzelnen Fragestellungen berichtet werden können, muss geklärt werden, ob die untersuchten Stichproben überhaupt repräsentativ sind. Diesem Aspekt widmet sich Kapitel 4.1. Anschliessend wird anhand der Darstellung und formalen Beschreibung der Zeitreihen beurteilt, ob das neue Auswertungsmodell die individuellen Unterschiede im Problemlöseverhalten beschreiben kann (Kapitel 4.2). Das dritte Kapitel präsentiert die Prozessprofile, die sich aufgrund inhaltlicher Überlegungen und aufgrund der Ergebnisse formaler Klassifizierungsmethoden ergeben. Die drei Fragestellungen zur Überprüfung des neuen Auswertungsmodells beantwortet Kapitel 4.4.

4.1 Repräsentativität der Teilstichproben

Für beide Teilstichproben wird geprüft, ob sie im Hinblick auf die für diese Arbeit interessierenden Merkmale repräsentativ sind. Dazu werden die beiden Stichproben hinsichtlich der 13 szenariospezifischen Leistungskennwerte (vgl. Kapitel 3.1.3.2) sowie weiterer Merkmale der Problembearbeitung (vgl. Tabelle 11 in Anhang D) mit der Gesamtstichprobe verglichen. Mittels *Kolmogorov-Smirnov-Test (KS-Test)* wird die Hypothese geprüft, dass die jeweilige Teilstichprobe und die Gesamtstichprobe aus identisch verteilten Populationen stammen. Der KS-Test spricht auf Verteilungsunterschiede aller Art an (zentrale Tendenz, Median, Dispersion, Schiefe, Exzess etc.) und ist der schärfste zur Zeit verfügbare Test zur Prüfung der genannten Hypothese (Bortz, Lienert & Boehnke, 2000, S. 299). Da für diese Vergleiche die Nullhypothese die „Wunschhypothese“ ist, wird ein *Signifikanzniveau von 25%* gewählt, um dadurch indirekt die Wahrscheinlichkeit eines β -Fehlers klein zu halten.

Zeigt der KS-Test ein signifikantes Resultat an, so wird mittels *U-Test nach Mann-Whitney* geprüft, ob sich die beiden Stichproben auch im mittleren Rang voneinander unterscheiden. Hierfür wird ein Signifikanzniveau von 5% bei zweiseitiger Hypothesenprüfung gewählt.

Für fast alle der getesteten Merkmale kann die Hypothese, dass die Verteilung in den Teilstichproben und der Gesamtstichprobe die gleiche ist, angenommen werden ($p > 25\%$). Die Überschreitungswahrscheinlichkeiten der jeweiligen Prüfgrößen liegen mehrheitlich sogar zwischen 60 bis 100%. Detaillierte Angaben zu allen gerechneten Einzelvergleichen sowie zu den Verteilungen der 13 Leistungskennwerte in den beiden Teilstichproben und der Gesamtstichprobe sind in Anhang D wiedergegeben (Tabelle 11 und Abbildung 17).

Für die Teilstichprobe 1 zeigen sich in Bezug auf die *Leistungskennwerte* keine signifikanten Verteilungsunterschiede, für die Teilstichprobe 2 ergibt sich lediglich für den Kennwert „Eingegangene Informationen effizient abrufen (p_{ow_m})“ ein signifikanter Verteilungsunterschied ($KS-Z = 1.07$, $p = .20$). Der diesbezügliche Unterschied im mittleren Rang ist jedoch klar nicht signifikant ($p > 5\%$). Hinsichtlich weiterer *Merkmale der Problembearbeitung* ergeben sich für die Teilstichprobe 1 in fünf Merkmalen und für die Teilstichprobe 2 in zwei Merkmalen nicht auszuschliessende Verteilungsunterschiede:

- Anzahl Videoeinsätze während Testphase (1: $KS-Z = 1.13$, $p = .15$; 2: $KS-Z = 1.15$, $p = .14$)
- Anzahl Videoeinsätze mit Meeresgrund während Testphase (1: $KS-Z = 1.09$, $p = .18$; 2: $KS-Z = 1.14$, $p = .15$)
- Anzahl Videoeinsätze mit Metall im Scanbereich während Testphase (1: $KS-Z = 1.42$, $p = .04$)
- Lesezeit im Hilfesystem während Training (1: $KS-Z = 1.21$, $p = .11$)
- Prozentualer Anteil der Lesezeit im Hilfesystem während Training (1: $KS-Z = 1.24$, $p = .09$)

Die ersten drei und die letzten zwei der aufgeführten Merkmale sind voneinander abhängig und beschreiben daher die gleichen Verhaltensweisen (Videoeinsätze während der Testphase, Lesen im Hilfesystem während dem Training). Diesbezügliche Einzelvergleiche hinsichtlich der zentralen Tendenz ergeben, dass die *Teilstichprobe 1* im Vergleich zur Gesamtstichprobe *während dem Training* prozentual mehr Zeit für das Lesen im Hilfesystem aufwendet ($U = 3200.0$, $Z = -1.97$, $p = .05$) und *während der*

Testphase weniger Videoeinsätze tätig (gesamthaft: $U = 3040.0$, $Z = -2.33$, $p = .02$; Videoeinsätze mit Metall im Videosichtbereich: $U = 3104.0$, $Z = -2.23$, $p = .03$). Für die Teilstichprobe 2 ergeben sich keine signifikanten Unterschiede hinsichtlich der zentralen Tendenz ($p > 5\%$).

Zusammenfassend ergibt sich, dass sich beide Teilstichproben im Hinblick auf die in dieser Arbeit interessierenden Merkmale nicht wesentlich von der Gesamtstichprobe unterscheiden. Von insgesamt 47 geprüften Merkmalen ergeben sich nur in *drei* (Teilstichprobe 2) beziehungsweise *fünf* (Teilstichprobe 1) Merkmalen signifikante Verteilungsunterschiede, wobei nur für drei dieser Merkmale (in der Teilstichprobe 1) auch ein Unterschied im mittleren Rang vorliegt. Damit können beide Teilstichproben als repräsentativ betrachtet werden.

Die folgenden Kapitel über die Prozessdarstellung und -beschreibung (Kapitel 4.2), die Definition von Prozessprofilen mittels Klassifizierung (Kapitel 4.3) sowie die Überprüfung des Auswertungsmodells anhand der szenariospezifischen Auswertung (Kapitel 4.4.1) und anhand des Expertenurteils (Kapitel 4.4.2) betreffen nur die Teilstichprobe 1. An dieser wurde der neue Ansatz entwickelt und das Beurteilungsschema generiert. Teilstichprobe 2 diente lediglich der Überprüfung des neuen Ansatzes an einem unabhängigen Datensatz (Kapitel 4.4.3).

4.2 Darstellung und Beschreibung der Zeitreihen

Für die graphische Darstellung der zeitlichen Entwicklung des Informationseröffnenden und Adäquaten Handelns wird die Achsenskalierung so gewählt, dass alle Zeitreihen – abgesehen von extrem abfallenden Verläufen – ganz abgebildet werden. Abbildung 4 zeigt am Beispiel einer durchschnittlich langen Problembearbeitung, wie die Zeitreihen für diese Arbeit aufbereitet werden. Neben den Zeitreihen für das Informationseröffnende und Adäquate Handeln (wie bisher jeweils abgekürzt mit Info und Ad) sind auch die *Systemereignisse* und die vom Probanden verfassten *Texteinheiten* eingetragen (vgl. gelbe Marken in Abbildung 4). Die Abkürzungen für die

Systemereignisse entsprechen den in Kapitel 3.2.2 beschriebenen (s. Tabelle 4, S. 83), wobei bei eingehenden Meldungen der Lesezeitpunkt und nicht der Sendezeitpunkt markiert ist. Für die vom Probanden erstellten Texteinheiten wird angegeben, welches Protokoll (p) oder welche Meldung (m) der Proband zu gegebenem Zeitpunkt eröffnet hat. Wurde ein Protokoll nicht gespeichert beziehungsweise eine Meldung nicht verschickt, so ist dies mit „n“ gekennzeichnet (z.B. p1n, wenn das erste Protokoll nicht gespeichert wurde). Diese zusätzlichen Angaben zum Spielverlauf sollen helfen, den Kurvenverlauf besser zu interpretieren.

Die auf diese Weise dargestellten Zeitreihen werden in den folgenden beiden Kapiteln qualitativ und quantitativ beschrieben.

4.2.1 Qualitative Beschreibung

Eine erste Sichtung der 35 Graphiken der Teilstichprobe 1 zeigt rasch, dass der neue Auswertungsansatz die individuellen Unterschiede im Problemlöseverhalten aufzeigen kann (Fragestellung 1). Abbildung 5 zeigt exemplarisch acht ausgewählte, sehr unterschiedliche Problembearbeitungen, auf die im Folgenden Bezug genommen wird. Für interessierte Leser sind alle anderen Graphiken in Anhang D abgebildet (Abbildung 18). Aus Platzgründen sind die Graphiken bei dieser Darstellung auf ein Minimum reduziert, so dass wohl die Zeitreihen, nicht aber die dazugehörigen Systemereignisse und vom Probanden erstellten Texteinheiten (gelbe Marken) gut erkenntlich sind.

Die Beispiele oben links (A) und rechts unten (H) in Abbildung 5 bilden die Extreme einer grossen Bandbreite von Problembearbeitungen. Das Vorgehen von A ist optimal: Er verschafft sich fortlaufend und effizient neue Informationen und verhält sich im aktuellen Handlungskontext adäquat. Kandidat H hingegen macht nur wenig informationseröffnende Schritte, er sammelt während des gesamten Spiels weniger Punkte als A während der Trainingsaufgabe. Zudem ist sein Handeln mehrheitlich nicht sinnvoll.

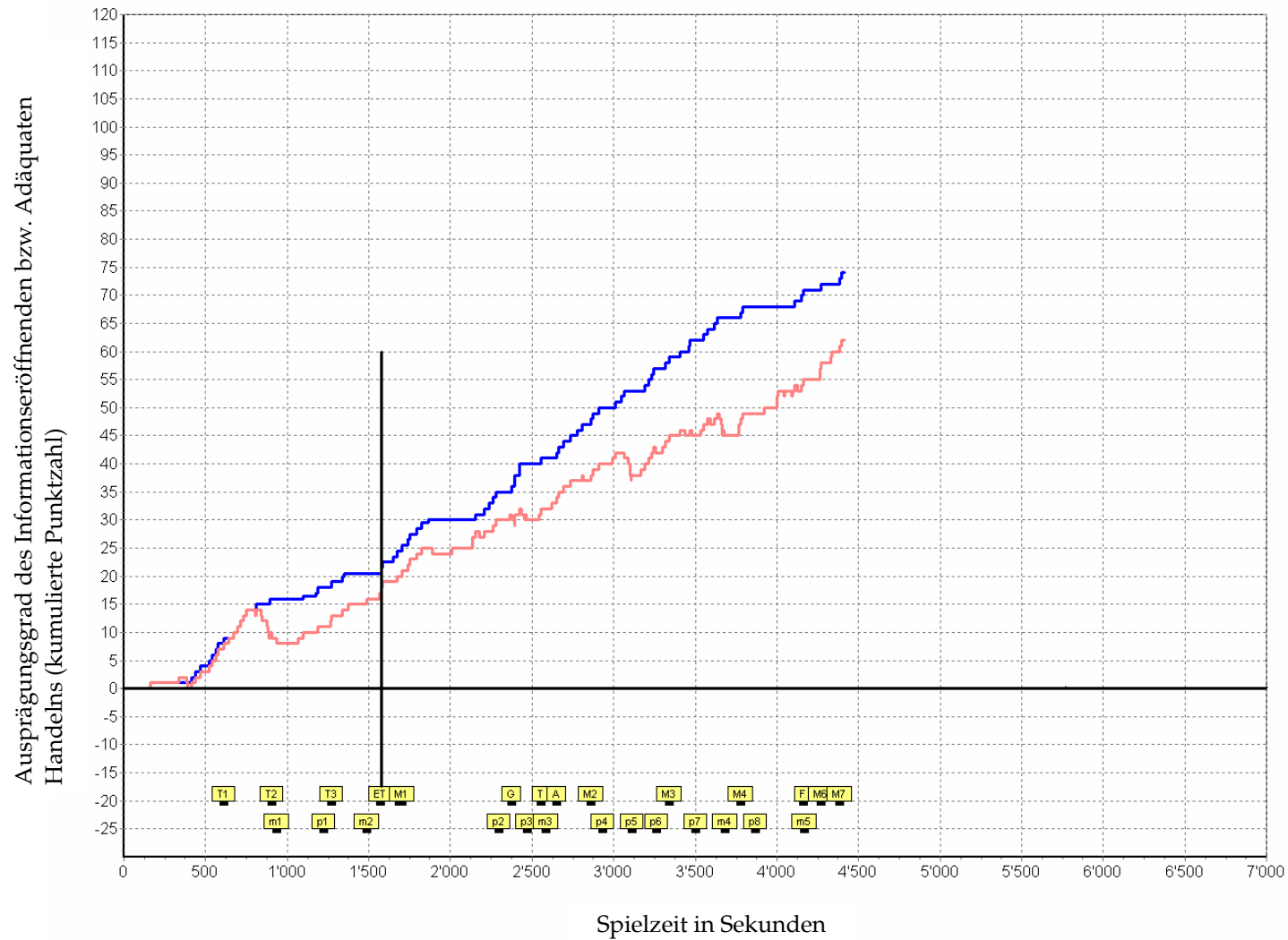


Abbildung 4: Darstellung der zeitlichen Entwicklung des Informationseröffnenden Handelns (blauer Graph) und des Adäquaten Handelns (hellroter Graph). Die gelben Marken kennzeichnen Systemereignisse und vom Probanden verfasste Texteinheiten (Abkürzungen s. Text). Die schwarze vertikale Linie markiert das Ende der Trainingsaufgabe.

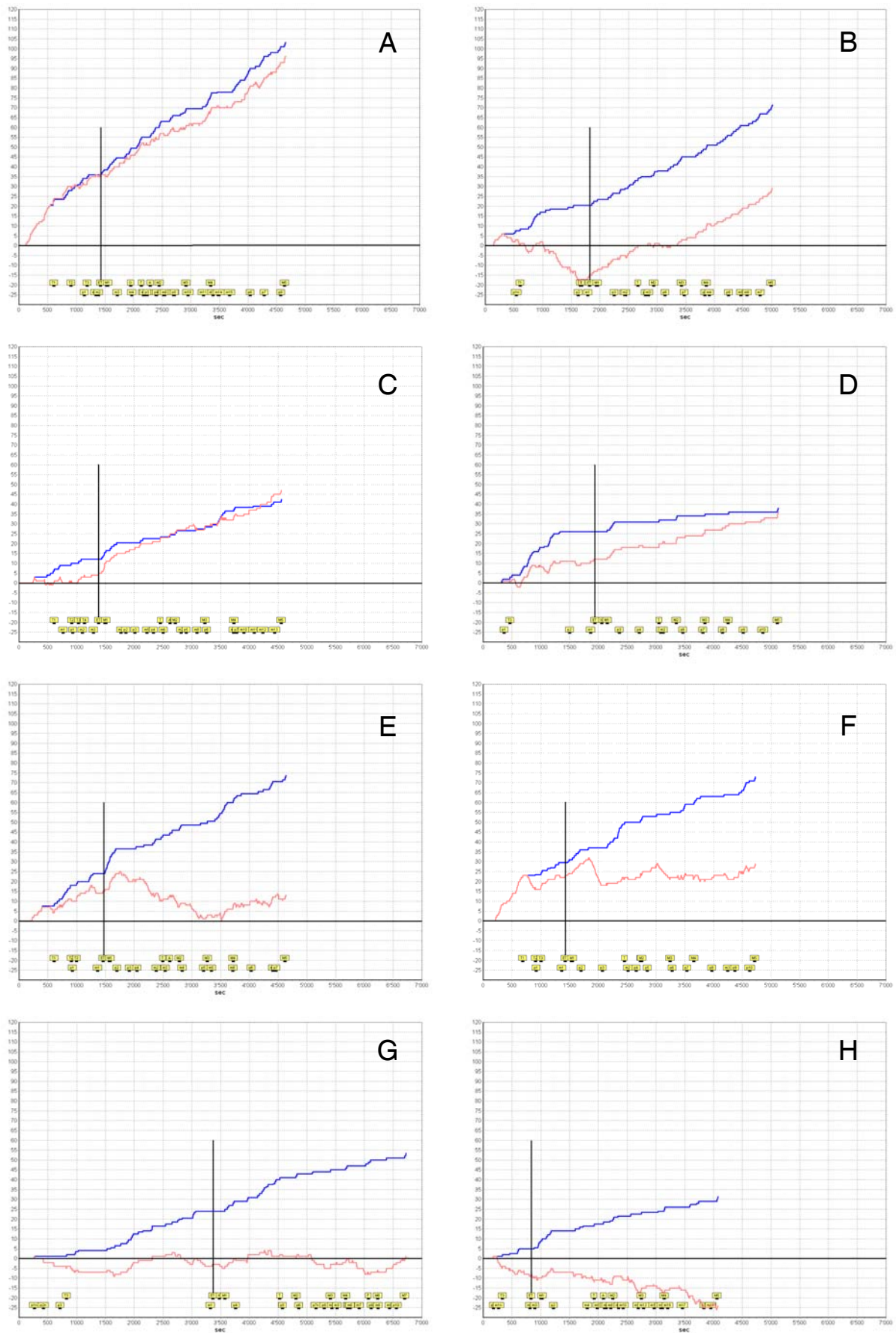


Abbildung 5: Beispiele unterschiedlicher Problembearbeitungen. Die Beispiele werden im Text näher erläutert.

Ein auffallendes Merkmal im Gesamtbild ist das *Auseinanderklaffen* der beiden Kurven, das bei fast allen der untersuchten Problembearbeitungen durch höhere Ausprägungsgrade für das Informationseröffnende Handeln zustande kommt. Nur bei einem einzigen Probanden liegt die Kurve für das Adäquate Handeln während mehr als der Hälfte der Bearbeitungszeit über der Info-Kurve. Das Auseinanderklaffen beziehungsweise der Abstand zwischen den beiden Kurven sagt etwas darüber aus, wie gut die (möglicherweise) aufgenommene Information umgesetzt wird. Für eine korrekte Interpretation muss allerdings der *Spielverlauf* mitberücksichtigt werden. So gibt es Probanden, deren Kurven nur während der Trainingsaufgabe auseinander laufen, während der Testphase jedoch mehr oder weniger parallel verlaufen (Bsp. B in Abbildung 5).

Werden nur die *Verläufe für die Testphase* betrachtet (durch Rekalibrierung der Kurven bei Trainingsende, d.h. Nullsetzung beim Zeitpunkt X oder ET), so verringert sich bei der Mehrheit der Probanden das Auseinanderklaffen, bei einigen wenigen wird der Abstand vergrößert (Bsp. C: Die Kurven sind nicht mehr mehrheitlich deckungsgleich) oder verändert (Bsp. D: Überschiessen der Info-Kurve wechselt zu Überschiessen der Ad-Kurve). Daneben zeigt sich, dass die im Vergleich zu anderen Probanden deutlich längere Spieldauer von Proband G vor allem auf eine *lange Trainingszeit* zurückzuführen ist, während die Zeit für die Testphase in etwa gleich ist wie bei den anderen.

Die separate Betrachtung der Verläufe für die Testphase hat auch einen Einfluss auf die *inhaltliche Bedeutungszuweisung* der einzelnen Kurven. Beispielsweise verläuft die (bei Trainingsende) rekalibrierte Kurve für das Adäquate Handeln von Proband E mehrheitlich im negativen und nicht mehr im positiven Bereich. Dieser Aspekt, das heisst die separate Betrachtung der beiden Spielphasen, wird bei der Definition der Prozessprofile in Kapitel 4.3 wieder aufgenommen.

An dieser Stelle folgen – als Voraussetzung für die Klassifizierung der Zeitreihen – einige Erläuterungen zu typischen Verlaufsformen und auffälligen Merkmalen der zeitlichen Entwicklung der beiden Konstrukte über die gesamte Bearbeitungszeit.

Wie erwartet zeigt sich in den Verläufen für das *Informationseröffnende Handeln* weniger Varianz als für das *Adäquate Handeln*. Die Kurven sind mehr oder weniger steil ansteigend und enthalten mehr oder weniger viele Plateaus (= Abschnitte ohne Steigung) unterschiedlicher Länge. Der in Abbildung 5 dargestellte Verlauf D beispielsweise weist während der Testphase fünf lange Plateaus auf, der Informationszuwachs in dieser Zeit ist minimal. Dieser Verlauf ist ein Beispiel für einen degressiv steigenden Verlauf (= abflachender Verlauf). Eindeutig progressiv steigende Verläufe können nicht beobachtet werden (auch in der Teilstichprobe 2 nicht).

Auch Nullverläufe (= kein Informationszuwachs) kommen (ebenfalls in beiden Teilstichproben) nicht vor. Auch schlechte Problemlöser erreichen bis zum Spielende mindestens 30 Punkte. Das hat damit zu tun, dass allein das *Lesen aller eingehenden Nachrichten* relativ viele Punkte gibt (17-18 bei einmaligem Lesen der Seiten und über 30 bei wiederholtem Lesen). Dass sich die (potentielle) Informationsaufnahme vor allem auf das Lesen von eingehenden Nachrichten beschränkt, ist daran zu erkennen, dass unmittelbar nach den entsprechenden Systemereignissen (T3, T, A, M1-M7) die Kurve ansteigt („Treppen“). Daneben kommen auch Probanden, die viel in der Bearbeitungshilfe oder in der Datenbank Nauti-Info lesen, auf eine hohe Punktzahl. Proband D beispielsweise ist jemand, dessen Informationsaufnahme sich nur auf das Lesen von Informationen beschränkt. Er macht während der gesamten Bearbeitung nur eine einzige Sonaraufnahme, die erst noch über dem Meeresgrund erfolgte (> 20 Meter über Grund).

Die Verläufe für das *Adäquate Handeln* sind vielfältiger und „unruhiger“. Neben mehr oder weniger steil (und mehr oder weniger konstant) ansteigenden Kurven (Bsp. A, D) zeigen sich auch abfallende Verläufe (Bsp. H) oder mehr oder weniger konstant um Null schwankende Verläufe (Bsp. G). Wie beim Informationseröffnenden Handeln kommen degressiv steigende Verläufe vor (Bsp. F: mehr oder weniger konstanter Anstieg im Training, während Testphase um den Trainingsendwert schwankend). Auch hier können keine eindeutig progressiv ansteigenden Kurven beobachtet werden.

Daneben kommen auch Mischformen vor, das heisst wechselnde Verläufe (Bsp. E: steigend, fallend, steigend). Einige Probanden zeigen den Verlaufswechsel nur beim

Übergang vom Training zur Testphase: Ein im Training um Null schwankender (Bsp. C) oder gar abfallender Verlauf (Bsp. B) mündet in einen konstant ansteigenden Verlauf während der Testphase.

Auffällige Merkmale sind die Häufigkeit und die Höhe der Ausschläge der Kurven („Spitzen“, Wechsel im Verlauf). Viele grössere Ausschläge geben einen Hinweis auf wenig konsistentes Verhalten, der Proband wechselt zwischen sinnvollen und nicht sinnvollen Arbeitsabschnitten (Bsp. F). Kleinere „Spitzen“ und Unregelmässigkeiten im Verlauf weisen auch die steil ansteigenden Kurven auf (Bsp. A), das heisst auch sehr gute Problemlöser machen ab und zu sinnlose Schritte.

4.2.2 Quantitative Beschreibung

Die qualitative Beschreibung über die Varianz in den Zeitreihen kann durch quantitative Angaben bestätigt werden. Die Ausprägungsgrade der beiden Konstrukte in den sieben Spielzeitpunkten sind normalverteilt ($.34 < KS-Z < .87$; $.43 < p < 1.0$). Die Varianz der Daten nimmt mit dem Spielverlauf zu, wobei dieser Anstieg für das Adäquate Handeln grösser ist (vgl. Abbildung 6). Die Endwerte für das Informationseröffnende Handeln variieren zwischen 31.5 und 117 ($M = 65.6$, $SD = 21.9$), die Endwerte für das Adäquate Handeln zwischen -34 und 106 ($M = 32.8$, $SD = 37.6$). Genauere Angaben über die Verteilungskennwerte in den verschiedenen Spielzeitpunkten sind Anhang D zu entnehmen (Tabelle 12 und Tabelle 13).

Bei beiden Konstrukten korrelieren die verschiedenen Ereignispunkte untereinander stark und hoch signifikant (vgl. Tabelle 14 in Anhang D). Die Korrelationen der einzelnen Zeitpunkte mit dem Endwert liegen zwischen .72 und .98 für das Informationseröffnende und zwischen .70 und .96 für das Adäquate Handeln. Auch der Zusammenhang zwischen Informationseröffnendem und Adäquaten Handeln ist in allen Spielzeitpunkten hoch signifikant. Die Korrelationen liegen zwischen .68 (im Zeitpunkt T3) und .82 (bei Spielende).

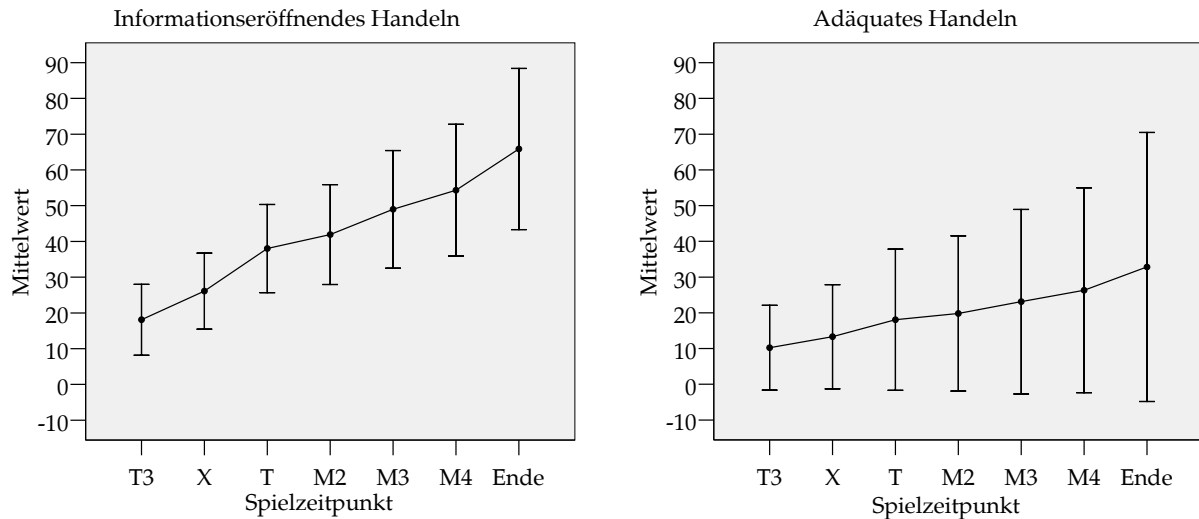


Abbildung 6: Entwicklung des Informationseröffnenden und Adäquaten Handelns in der Teilstichprobe 1. Angegeben sind die Mittelwerte und Standardabweichungen (Balken) zu ausgewählten Spielzeitpunkten ($n = 35$, für T3 und T ist $n = 34$).

Wie erwartet variiert die *Spieldauer* der 35 Probanden beträchtlich: Die kürzeste Bearbeitung dauerte knapp eine Stunde (62 Minuten), die längste knapp zwei Stunden (112 Minuten; $M = 78$ Minuten, $SD = 11$ Minuten). Wie Abbildung 7 zeigt, ist diese Variabilität vor allem auf *unterschiedlich lange Trainingszeiten* zurückzuführen: Die Werte für die Trainingszeit streuen zwischen 17 und 58 Minuten ($M = 29$ Minuten, $SD = 9$ Minuten), die Werte für die Testzeit nur zwischen 38 und 55 Minuten ($M_d = 51$ Minuten). Zudem korreliert die Gesamtspielzeit mit der Trainingszeit um $r = .94$ ($p < .001$), mit der Testzeit jedoch nur um $r_s = .58$ ($p < .001$). Der Zusammenhang zwischen Trainings- und Testzeit ist nicht signifikant.

Die grosse Variabilität in der Trainingszeit ist zum Teil darauf zurückzuführen, dass der Spielverlauf während dem Training weniger standardisiert ist und damit einen grösseren Handlungsspielraum zulässt als während der Testphase. In der Testphase erhalten alle Probanden im Abstand von wenigen Minuten immer wieder bestimmte Meldungen von der Einsatzleitung, unabhängig von ihrem Verhalten (mit Ausnahme der Meldung A). In der Trainingsphase sind Sendezeitpunkt sowie Inhalt der Meldungen, die der Proband erhält, jedoch stark von seinem Verhalten abhängig (vgl. Kapitel 3.1.2). Sobald der Proband ein erstes Protokoll erstellt hat, wird er

aufgefordert, das Training (durch Auftanken am Mutterschiff) zu beenden – allerdings unabhängig vom Inhalt dieses Protokolls und unabhängig davon, ob er bereits erste Erfahrungen im Umgang mit Nautilus gesammelt hat oder nicht. Leistet er der Aufforderung Folge, so ist das Training beendet und die Testphase beginnt. Ignoriert der Proband die Aufforderung, so kann er weiter spielen, bis die Energiereserve von Nautilus aufgebraucht ist.

In Abhängigkeit von der Reaktion des Probanden auf die Meldungen sind damit unterschiedlich lange Trainingszeiten möglich. Das bedeutet auch: Die Probanden haben nicht unbedingt die gleichen Möglichkeiten, sich während der Trainingsphase problemrelevante Informationen zu verschaffen oder umgekehrt, sich „falsch zu verhalten“. Dies erschwert eine vergleichende Interpretation der Verläufe während dieser Spielphase. Ausserdem führt der grössere Handlungsspielraum im Training auch zu einer grösseren Vielfalt an Verlaufsformen für das Informationseröffnende und Adäquate Handeln während dieser Spielphase. Beides muss für die Definition der Prozessprofile berücksichtigt werden.

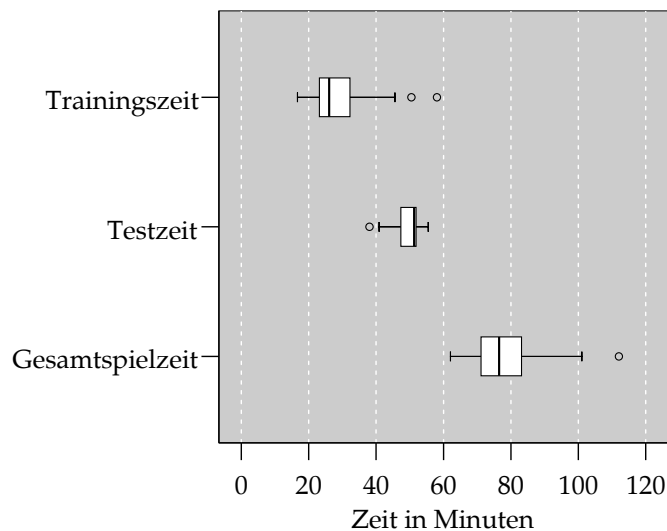


Abbildung 7: Verteilung der Spielzeit in der Teilstichprobe 1.

4.3 Definition von Prozessprofilen mittels Klassifizierung der Zeitreihen

Die Ergebnisse und Überlegungen aus dem vorangegangenen Kapitel bilden die Grundlage für die Definition der Prozessprofile. Die Ausführungen haben gezeigt, dass aufgrund der grossen Varianz im Training (hinsichtlich Handlungsspielraum, Trainingszeit und Verlaufsformen) neben dem Gesamtverlauf der beiden Konstrukte auch die beiden Spielphasen Training und Testphase getrennt betrachtet werden müssen. Dabei soll der Testphase mehr Gewicht zukommen, einmal aufgrund der längeren Dauer (gemäss Instruktion: 15 Minuten Trainingszeit und ca. 50 Minuten Testzeit), einmal weil diese auch für den Probanden als Ernstfall deklariert wird, während die Trainingsaufgabe „nur“ zum Üben dient.

Kapitel 4.3.1 erläutert, welche Prozessprofile und Klassifikationsregeln definiert wurden (Fragestellung 2). Anschliessend wird überprüft, ob sich die inhaltlich orientierte Klassifizierung durch formale Methoden (Clusteranalysen, Multidimensionale Skalierung) bestätigen lässt (Kapitel 4.3.2, Fragestellung 3). Kapitel 4.3.3 beinhaltet das aus den Prozessprofilen resultierende Beurteilungsschema und zeigt, wie sich die Probanden der Teilstichprobe 1 auf die verschiedenen Problemlösetypen verteilen.

4.3.1 Klassifizierung aufgrund inhaltlicher Überlegungen

Zuerst werden die Prozessprofile für das Informationseröffnende Handeln definiert (Kapitel 4.3.1.1), anschliessend jene für das Adäquate Handeln (Kapitel 4.3.1.2). Die Prozessprofile sind, wie bei der Beschreibung des Auswertungsmodells bereits erläutert (vgl. Kapitel 2.3), hierarchisch aufsteigend geordnet, das heisst Profil 1 beschreibt jeweils das beste Vorgehen. Entsprechend sind bei der Beschreibung der Klassifikationsregeln nur die Kriterien aufgeführt, die die jeweilige Klasse von der hierarchisch tieferen Klasse abgrenzen.

4.3.1.1 Prozessprofile für das Informationseröffnende Handeln

Wie in Kapitel 4.2 dargelegt, zeigt sich in den Verläufen für das Informationseröffnende Handeln weniger Varianz als in den Verläufen für das Adäquate Handeln. Es lassen sich daher weniger klar voneinander abgrenzbare Gruppen definieren, die Gruppengrenzen sind weniger eindeutig und mögen zu Recht etwas willkürlich erscheinen. Mit der Wahl von *drei Klassen* gelingt die Gruppeneinteilung aber dennoch recht gut. Die Verläufe der Gruppen 2 und 3 sind dabei deutlich heterogener als die Verläufe der Gruppe 1, und zudem ist Gruppe 1 nur halb so gross wie die anderen beiden Gruppen ($n_1 = 7$, $n_2 = 13$, $n_3 = 15$).

Entscheidende Kriterien für die Klassifizierung sind die Verlaufsform, die absolute Informationsmenge, die sich der Proband während der Problembearbeitung eröffnet, die Relation zwischen der Informationsmenge während der Trainings- und der Testphase, sowie die Zeit, die der Proband für das Schaffen der Informationsmenge benötigt.

Die drei Prozessprofile werden im Folgenden zusammenfassend beschrieben. Da die Verlaufsformen, die die Prozessprofile beinhalten, sehr einfach zu beschreiben sind, wird auf eine graphische Darstellung verzichtet. Stattdessen wird jeweils auf die in Kapitel 4.2.1 diskutierten Beispiele aus Abbildung 5 verwiesen (s. Seite 98).

Prozessprofil 1

Verlaufsform	Der Verlauf ist sehr steil und konstant ansteigend, schon während der Trainingsaufgabe (Bsp. A).
Zuordnungskriterien	Vor dem Zeitpunkt $t = 4000$ Sekunden beträgt der Ausprägungsgrad der Zeitreihe mindestens 80 Punkte (bzw. der Endwert beträgt mindestens 80 Punkte, falls die Spieldauer weniger als 4000 Sekunden beträgt). Während der Testphase werden mindestens 45 Punkte erreicht.
Interpretation	„Sehr gut“. Der Proband eröffnet sich mit fast jedem Handlungsschritt neue, problemrelevante Informationen und geht dabei sehr effizient vor. Neben dem Lesen von eingehenden Informationen, dem Abfragen von Informationen aus der Systemhilfe oder der Datenbank Nauti-Info verschafft sich der Proband vor allem durch korrektes Einsetzen der Such- und Prüfinstrumente neue Informationen.

Prozessprofil 2

Verlaufsform	Der Verlauf ist weniger steil und zum Teil weniger konstant ansteigend als bei Prozessprofil 1, aber ebenfalls schon im Training ansteigend (Bsp. B, E, F). Einige Verläufe weisen längere Plateaus auf.
--------------	--

Zuordnungskriterien	Der Endwert beträgt mindestens 45 Punkte und während der Testphase werden mindestens 30 Punkte erreicht.
Interpretation	„Gut“. Der Proband eröffnet sich von Anfang an konstant neue, problemrelevante Informationen, ist aber weniger effizient als Probanden aus Gruppe 1 (was häufig auf einen weniger erfolgreichen Einsatz der Such- und Prüfinstrumente zurückzuführen ist).
<i>Prozessprofil 3</i>	
Verlaufsform	Dieses Profil setzt sich aus 3 <i>Verlaufsformen</i> zusammen. Der Verlauf kann entweder <i>flach ansteigend</i> (Profil 3.1, Bsp. C, H) oder <i>degressiv ansteigend</i> (Profil 3.2, Bsp. D) sein, wobei bei letzterem ab Beginn der Testphase eine deutliche Abflachung erfolgt. Zudem zählen auch Verläufe mit sehr langer Trainingszeit zu diesem Profil, unabhängig von der Kurvenhöhe („ <i>lange Verläufe</i> “, Profil 3.3, Bsp. G).
Zuordnungskriterien	Der Endwert ist geringer als 45 Punkte, oder während der Testphase werden weniger als 30 Punkte erreicht, oder die Trainingszeit beträgt mehr als 2500 Sekunden (Profil 3.3).
Interpretation	„Mässig bis schlecht“. Der Proband eröffnet sich gesamthaft wenig neue, problemrelevante Informationen. Bei <i>Profil 3.1</i> nimmt der Proband zwar konstant, aber wenig effizient neue Informationen auf, so dass er im Vergleich zu Probanden der Profile 1 und 2 deutlich weniger Punkte erreicht. Meistens kann dies auf den nicht korrekten Einsatz von Sonar- und Videogerät zurückgeführt werden, wobei der Proband sich vorwiegend durch Lesen von (eingehenden) Informationen Punkte holt. Bei <i>Profil 3.2</i> beschränkt sich die potentielle Aufnahme neuer Information vorwiegend auf die Trainingsaufgabe (und dabei auf das Lesen in der Bedienungshilfe). Ein <i>Profil-3.3</i> -Proband eröffnet sich mehr oder weniger konstant und durch verschiedenste Aktionen neue Informationen, geht aber äusserst ineffizient vor, indem er für das Training zu viel Zeit beansprucht (> 40 Minuten). Obwohl bei diesem Verlauf auch Endwerte vorkommen können, die eine Einordnung in Profil 1 oder 2 zulassen, werden solche „ <i>langen Verläufe</i> “ immer Profil 3 zugeordnet. Die absolute Informationsmenge, die sich ein Proband eröffnet, muss auch in Relation zur dafür verwendeten Zeit beurteilt werden.

4.3.1.2 Prozessprofile für das Adäquate Handeln

Bei den Verläufen für das Adäquate Handeln können mehrere typische Verläufe identifiziert werden (vgl. Kapitel 4.2.1). Da eine Klassifizierung von etwa acht Gruppen für ein ökonomisches Beurteilungsschema wenig sinnvoll ist, werden diese zu *vier Gruppen* zusammengefasst, wobei die ersten drei Gruppen in etwa gleich gross sind und die schlechteste vierte Gruppe die grösste und heterogenste Gruppe darstellt ($n_1 = 6$, $n_2 = 7$, $n_3 = 7$, $n_4 = 15$).

Entscheidend für die Klassifizierung ist die Verlaufsform als Ganzes (steil ansteigend, flach ansteigend, um Null schwankend, abfallend) und die Relation zwischen Trainings- und Testphase, wobei dem Verlauf in der Testphase das höhere Gewicht zukommt (genauer: dem am Trainingsendwert rekalierten Verlauf).

Die vier Prozessprofile werden im Folgenden zusammenfassend beschrieben. Wie im vorangegangenen Kapitel wird auf eine graphische Darstellung der verschiedenen Verlaufsformen verzichtet, da diese sehr einfach zu beschreiben sind. Wo vorhanden, wird wiederum auf die in Kapitel 4.2.1 diskutierten Beispiele aus Abbildung 5 verwiesen (s. Seite 98).

Prozessprofil 1

Verlaufsform	Der Verlauf ist sehr steil und konstant ansteigend, schon während dem Training (Bsp. A).
Zuordnungskriterien	Der Endwert beträgt mindestens 80 Punkte und während der Testphase werden mindestens 50 Punkte erreicht. Falls die Gesamtspieldauer kleiner als 4000 Sekunden ist, genügt ein Endwert von mehr als 70 Punkten.
Interpretation	„Sehr gut“. Der Proband macht nur wenige nicht sinnvolle Schritte und verhält sich ansonsten äusserst adäquat. Er setzt die verschiedenen Instrumente (Funktionseinheiten des Steuerungscomputers) korrekt und sinnvoll ein. Das gezeigte Problemlöseverhalten ist optimal.

Prozessprofil 2

Verlaufsform	Dieses Profil beschreibt <i>zwei Verlaufsformen</i> . Der Verlauf ist ebenfalls (sowohl während dem Training wie auch der Testphase) <i>steil ansteigend</i> (Profil 2.1), jedoch ist der Anstieg weniger steil und weniger konstant wie bei Prozessprofil 1. Oder der Verlauf ist <i>nur während der Testphase steil ansteigend</i> (Profil 2.2), während dem Training ist der Verlauf flach oder sogar abfallend (so dass der Endwert geringer ausfällt als bei Prozessprofil 1, Bsp. B, C).
Zuordnungskriterien	Der Endwert beträgt mindestens 50 Punkte oder während der Testphase werden mindestens 30 Punkte erreicht.
Interpretation	„Gut“. Der Proband verhält sich adäquat, wobei das sinnvolle Vorgehen vereinzelt durch kurze Phasen mit wenig oder nicht sinnvollem Verhalten unterbrochen wird. Ein <i>Profil-2.2</i> -Proband hat Startschwierigkeiten, das heisst während dem Training verhält er sich wenig bis nicht sinnvoll, zeigt dann aber während der Testphase adäquates Verhalten. Dieses Verhalten könnte darauf zurückzuführen sein, dass der Proband während dem Training ausprobiert und übt, und erst während der Testphase, im Ernstfall und unter Druck, Leistung erbringt.

Prozessprofil 3

Verlaufsform	Auch dieses Profil besteht aus <i>zwei Verlaufsformen</i> . Der Verlauf ist (während dem Training und der Testphase) mehr oder weniger konstant und <i>flach ansteigend</i> (Profil 3.1, Bsp. D) oder eine <i>Mischform</i> (Profil 3.2). Mischformen sind wechselnde Verläufe (z.B. kurzer Anstieg, rasches Abfallen, danach konstanter langsamer Anstieg). Allen Verläufen gemeinsam ist, dass der Verlauf während der Testphase <i>gesamthft tendenziell steigend</i> ist und dass praktisch keine negativen Werte vorkommen.
Zuordnungskriterien	Der Endwert beträgt mindestens 20 Punkte und ist grösser als der Trainingsendwert zum Zeitpunkt X. Während der Testphase werden mindestens zehn Punkte erreicht.
Interpretation	„Mässig“. Der Proband verhält sich mässig adäquat. Bestimmte Aspekte der Problemlösung scheint er verstanden zu haben, andere jedoch nicht (häufig zu beobachten ist ein korrekter Einsatz des Sonargerätes bei mehrheitlich falschem Einsatz des Videogerätes).

Prozessprofil 4

Verlaufsform	Zu diesem Profil gehören <i>vier Verlaufsformen</i> . Am häufigsten sind die mehrheitlich <i>um Null schwankenden</i> (Profil 4.1, Bsp. G) und <i>abfallenden Verläufe</i> (Profil 4.2, Bsp. H). Daneben werden aber auch <i>degressiv ansteigende</i> Verläufe, deren Werte während der Testphase mehr oder weniger um den Trainingsendwert schwanken, diesem Profil zugeordnet (Profil 4.3, Bsp. F). Ebenfalls in diese Gruppe fallen verschiedene <i>Mischformen</i> (Profil 4.4, Bsp. E), die die Kriterien für Profil 3 nicht erfüllen (z.B. ist der Endwert trotz tendenziellem Anstieg während der Testphase geringer als 20 Punkte).
Zuordnungskriterien	Der Endwert ist kleiner als 20 Punkte, oder während der Testphase werden weniger als zehn (Plus-) Punkte erreicht.
Interpretation	„Schlecht“. Das Verhalten des Probanden ist mehrheitlich wenig bis nicht sinnvoll. Wesentliche Merkmale der Problemlösung bleiben dem Probanden verborgen. Diese Nichteinsicht in die Problemlösung kann nur Teilsysteme betreffen (z.B. nur die Steuerung), oder – bei den abfallenden Verläufen, das heisst <i>Profil 4.2</i> – alle zur Verfügung stehenden Instrumente. Bei einigen Probanden zeigt sich die schlechte Leistung erst mit zunehmend schwierig werdender Aufgabe (d.h. während der Testphase, <i>Profil 4.3</i>).

Als nächstes interessiert nun die Frage, ob die in diesem Kapitel vorgeschlagene Klassifizierung der Info- und Ad-Zeitreihen durch formale Klassifizierungsmethoden bestätigt werden kann.

4.3.2 Klassifizierung aufgrund formaler Methoden

Analog zum vorhergehenden Kapitel wird auch hier zuerst die Lösung für das Informationseröffnende Handeln präsentiert (Kapitel 4.3.2.2), anschliessend jene für das Adäquate Handeln (Kapitel 4.3.2.3). Dabei werden jeweils zuerst die Ergebnisse der Klassifizierung anhand der Ereignispunkte dargestellt *und interpretiert*, danach die Ergebnisse der Klassifizierung anhand der Euklidischen Distanzen. Eine zusammenfassende Diskussion der verschiedenen Ergebnisse folgt in Kapitel 4.3.2.4.

Vor der Ergebnisdarstellung sollen im Folgenden einige Aspekte der dabei verwendeten Methoden (Clusteranalyse, Multidimensionale Skalierung) erläutert werden.

4.3.2.1 Anmerkungen zu den verwendeten Methoden

Mit der *Clusteranalyse* können Objekte so gruppiert werden, dass die Unterschiede zwischen den Objekten einer Gruppe möglichst gering und die Unterschiede zwischen den Objekten verschiedener Gruppen möglichst gross sind. Das Ergebnis einer Clusteranalyse ist insbesondere von der Definition der Ähnlichkeit der Objekte und der Wahl des Fusionierungskriteriums abhängig. In dieser Arbeit dient die Clusteranalyse dazu, eine nach inhaltlichen Überlegungen vorgeschlagene Klassifizierung von Zeitreihen auf Plausibilität zu überprüfen. Die Ähnlichkeit der Zeitreihen wird dabei einerseits beurteilt durch die Ausprägungsgrade der beiden Konstrukte zu ausgewählten Systemereignissen (Ereignispunkte), und andererseits durch die Kurvenabstände (formalisiert als die Euklidische Distanz zwischen je zwei Zeitreihen, vgl. Kapitel 3.2.3). Für alle Fusionierungsmethoden dient deshalb der *Quadrierte Euklidische Abstand* als Ähnlichkeitsmass.

Trotz einer Vielzahl von existierenden Clusteralgorithmen ist es bis heute nicht möglich, in einer vernünftigen Zeit die beste aller möglichen Clusterlösungen zu bestimmen (Bortz, 2005, S. 571). Nach Milligan (1981; zit. nach Bortz, 2005, S. 573) erzielt die Ward-Methode zumindest für Ähnlichkeitsmasse, die sich als Euklidische Distanzen interpretieren lassen, die besten Resultate. Allerdings kann sich der Ward-

Algorithmus nachteilig auswirken, wenn die bestmögliche Klassifizierung aus Clustern unterschiedlicher Grösse besteht, da der Algorithmus dazu tendiert, Unterschiede in den Gruppengrössen auszugleichen, je länger die Fusionierung anhält. Da die durch die inhaltlich orientierte Klassifizierung vorgeschlagenen Gruppen unterschiedlich gross sind, werden deshalb neben der *Ward-Methode* jeweils auch die *Average-linkage-Methoden* (Within-groups- und Between-groups-linkage) gerechnet. Diese sind im Falle ungleicher Clustergrössen dem Ward-Algorithmus überlegen (Bortz, 2005, S. 578). Die zusätzliche Verwendung der *Single-linkage-Methode* dient der Identifikation von Ausreissern. Da extreme Werte die Art der Clusterbildung beeinflussen können, werden *alle Berechnungen auch ohne Ausreisser* durchgeführt und geprüft, ob sich dadurch andere Clusterzugehörigkeiten ergeben.

Die mit diesen hierarchisch-agglomerativen Verfahren gefundenen Clusterlösungen werden anschliessend mit der nicht-hierarchischen *k-means-Methode* optimiert.

Das Ergebnis einer Clusteranalyse ist nicht nur von der Wahl des Ähnlichkeitsmasses und der Wahl des Clusteralgorithmus abhängig, sondern ganz wesentlich auch von der *Auswahl der Merkmale*, anhand derer die Ähnlichkeit der Objekte beschrieben wird. Mehrere, wechselseitig korrelierte Merkmale überbetonen bestimmte Aspekte und können zu Verzerrungen führen, wenn diese Übergewichtung der inhaltlichen Problemstellung nicht angepasst ist. Bei der Clusteranalyse über die Ereignispunkte trifft dies zu: Es werden mehrere Variablen berücksichtigt, die untereinander hoch korrelieren. Da die Variablen jedoch den gleichen Prozess beschreiben und nicht Merkmale darstellen, die verschiedene Eigenschaften abbilden, sollten die Interkorrelationen zwischen den Ereignispunkten nicht zu Verzerrungen führen, sondern im Gegenteil, sich positiv auf die Abbildung der Ähnlichkeit der Zeitreihen auswirken. Deshalb sollen *alle Ereignispunkte* berücksichtigt werden. Da in den Spielzeitpunkten T3 und T jeweils ein fehlender Wert vorkommt (vgl. Kapitel 4.2.2), werden die Clusteranalysen über folgende Ereignispunkte gerechnet: X, M2, M3, M4 und Ende. Beim Informationseröffnenden Handeln wird zudem die *Spieldauer* (Trainings- und Gesamtspielzeit) mitberücksichtigt, da diese auch in den Klassifikationsregeln verschiedentlich enthalten ist.

Weil die Durchführung der Clusteranalyse lediglich der Überprüfung einer vorgegeben Klassifizierung dient, wird auf analytische Abbruchkriterien zur Bestimmung der optimalen Clusteranzahl verzichtet. Stattdessen wird anhand der Entwicklung der Fehlerquadratsumme überprüft, ob die gewählte Clusteranzahl vertretbar ist. Dem Dendrogramm kann entnommen werden, welcher Fusionierungsschritt mit einem sprunghaften Fehlerquadratsummen-Zuwachs verbunden ist. Ebenfalls anhand des Dendrogramms wird beurteilt, ob die Gruppenzusammensetzung der gewählten Klassifizierung entspricht, das heisst, ob die „richtigen“ Probanden den entsprechenden Gruppen (Prozessprofilen) zugeteilt werden.

Auch die *Multidimensionale Skalierung* (MDS) dient der Überprüfung der vorgeschlagenen inhaltlich orientierten Klassifizierung und wird deshalb in einer etwas unüblichen Art und Weise angewendet. Es geht nicht darum, die Ähnlichkeiten zwischen den Zeitreihen durch wenige Dimensionen zu beschreiben und diesen eine Bedeutung zuzuweisen. Ziel ist, die „Grösse“ der Ähnlichkeiten zwischen den Zeitreihen zu *visualisieren*, das heisst aufzuzeigen, welche Zeitreihen (Probanden) nahe beieinander liegen und sich damit ähnlicher sind als andere. Während die Clusteranalyse vor allem Aussagen darüber macht, wie viele Gruppen gebildet werden sollen und aus welchen Probanden diese Gruppen bestehen, kann die MDS darstellen, wie ähnlich (bzw. unähnlich) sich die Probanden *innerhalb* dieser Gruppen sind und wie eindeutig oder unscharf die Gruppengrenzen sind. Anhand der ermittelten Konfiguration wird deshalb beurteilt, ob und wie gut die mit anderen Methoden ermittelten Lösungen (inhaltlich orientierte Klassifizierung, Clusteranalyse über Ereignispunkte, Clusteranalyse über Euklidische Distanzen) mit der Lösung der MDS übereinstimmen.

Da die Ähnlichkeiten zwischen den Zeitreihen über Euklidische Distanzen operationalisiert werden, kann eine metrische MDS (für eine zweidimensionale Ähnlichkeitsmatrix) gerechnet werden. Es werden ein zweidimensionaler Darstellungsraum und ein lineares Regressionsmodell gewählt. Die Güte der ermittelten Konfiguration wird durch die Kriterien *S-stress nach Young* und *Kruskal's Stress* sowie durch die *quadierte Korrelation zwischen den Ähnlichkeitsmassen und den Distanzen* (= RSQ) beurteilt.

Abschliessend noch einige Anmerkungen zu den Voraussetzungen der angewendeten Methoden beziehungsweise zur Datengrundlage: Die angewendeten Methoden zur Beurteilung der Ähnlichkeit der Zeitreihen setzen metrische Daten voraus. Mit der Formalisierung der Ähnlichkeit über die Ereignispunkte und über die Abstände zwischen den Zeitreihen in Form Euklidischer Distanzen liegen auch tatsächlich metrische Daten vor. Dennoch darf nicht vergessen werden, dass die ursprünglichen Daten den Anforderungen an eine metrische Skala nicht genügen: Die Zeitreihen basieren auf der Aufsummierung von Mikroverhaltensdaten, die höchstens Rangskalencharakter aufweisen. Sowohl die Berechnung der Euklidischen Distanzen wie auch die angewendeten Klassifizierungsmethoden sind daher etwas gewagt. Da es aber lediglich darum geht, die Plausibilität einer inhaltlich begründeten Klassifizierung zu überprüfen, scheint deren Anwendung doch vertretbar.

Im Folgenden werden die Ergebnisse der verschiedenen Berechnungen präsentiert. Für ein besseres Verständnis werden die Gruppen jeweils mit „Profilen“ bezeichnet, wenn damit die inhaltliche Klassifizierung gemeint ist, und mit „Clustern“, wenn damit das Ergebnis der Clusteranalyse oder der MDS gemeint ist.

4.3.2.2 Lösung und Interpretation für das Informationseröffnende Handeln

Die inhaltlich orientierte Klassifizierung lässt sich durch die *Klassifizierung anhand der Ereignispunkte* mehrheitlich bestätigen. Allerdings legen alle gerechneten Clusteranalyse-Varianten eine Lösung mit vier (statt drei) Clustern nahe. Dabei bilden vier Probanden des Profils 3 einen eigenen Cluster (Nr. 5, 7, 23 und 27). Die entsprechend der inhaltlich orientierten Klassifizierung besten drei Probanden aus Profil 1 bilden ebenfalls einen eigenen Cluster. Daneben bildet die Mehrheit der Probanden aus Profil 2 (zusammen mit vier Probanden aus Profil 1) einen eigenen Cluster, ebenso die Mehrheit der Probanden aus Profil 3 (zusammen mit – je nach Algorithmus – drei bis vier Probanden aus Profil 2). Abbildung 8 zeigt beispielhaft die Lösung mit dem Ward-Algorithmus.

Dendrogram using Ward Method

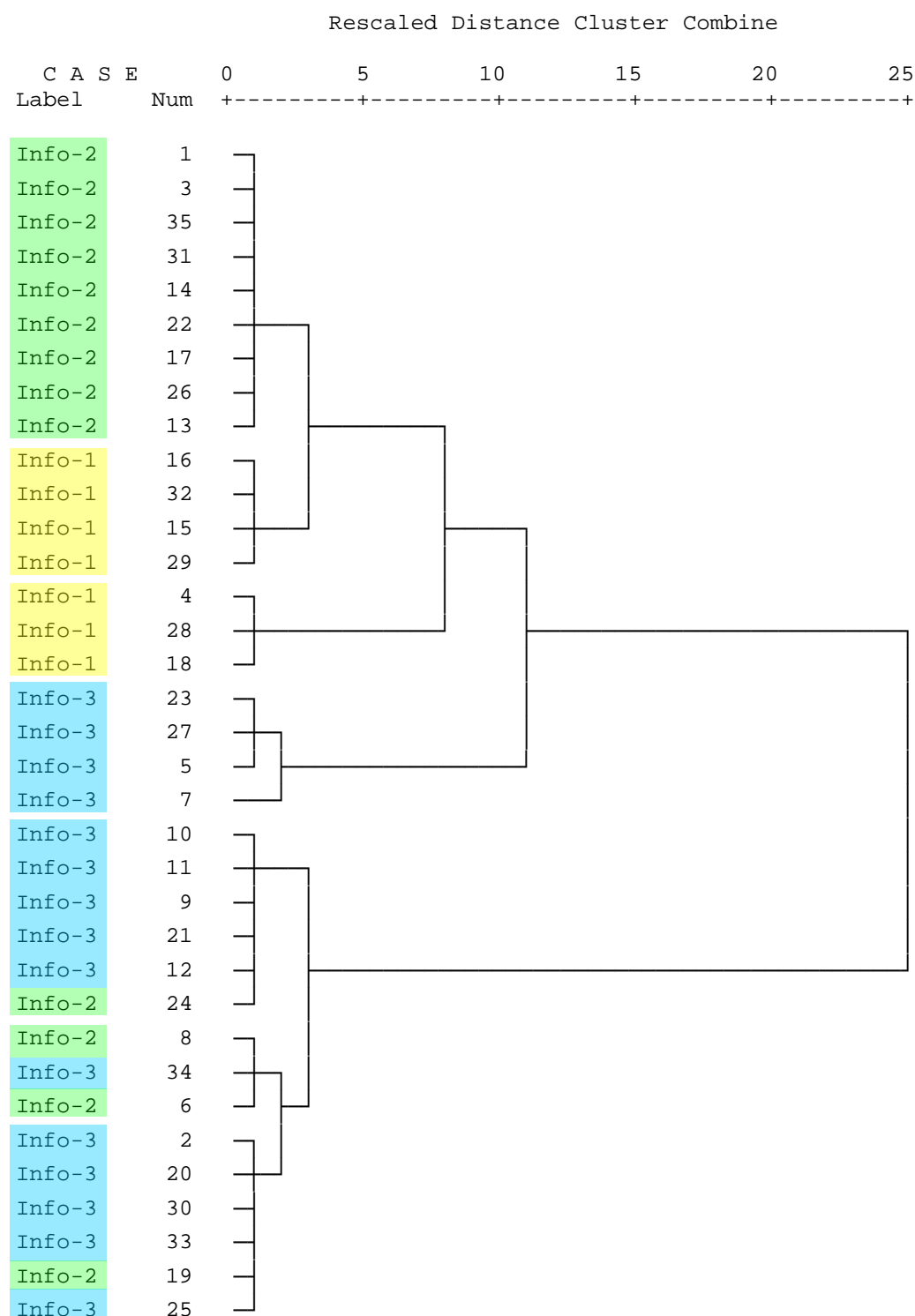


Abbildung 8: Dendrogramm der Clusteranalyse über die Ereignispunkte für das Informationseröffnende Handeln. Die Probanden sind durch die Gruppeneinteilung entsprechend der inhaltlich orientierten Klassifizierung gekennzeichnet, wobei gelb Profil 1, grün Profil 2 und blau Profil 3 markiert (n = 35).

Der Hauptunterschied zur inhaltlich orientierten Klassifizierung liegt in der Aufteilung der Profile 1 und 3 zu je zwei separaten Clustern. Diese Aufteilung zeigt sich bei genaueren Analysen allerdings nicht mehr. Anhand der Single-linkage-Methode lassen sich *vier Probanden als Ausreisser* identifizieren. Werden diese bei den Berechnungen ausgeschlossen, so führt dies (bei der Ward-Methode) zu einer *eindeutigen Drei-Clusterlösung*, wobei *alle Probanden aus Profil 1 einen eigenen Cluster bilden*. Daneben bilden die Probanden aus Profil 2 ebenfalls einen eigenen Cluster, wobei (analog zur Clusteranalyse über alle Probanden) drei bis vier Profil-2-Probanden in den Cluster 3 eingeordnet werden.

Die festgestellten Abweichungen von der inhaltlich orientierten Klassifizierung (vier Ausreisser, drei bis vier Probanden im Cluster 3 statt 2) lassen sich einfach erklären. Bei den vier *Ausreissern* handelt es sich um die vier Profil-3-Probanden, die bei der Clusteranalyse über alle Probanden einen separaten Cluster bilden (Nr. 5, 7, 23 und 27). Alle vier verfügen (als einzige) über sehr lange Trainingszeiten und damit *lange Verläufe* (Profil 3.3). Offenbar wird der Zeitaspekt durch die Berücksichtigung der Trainings- und der Gesamtspielzeit als weitere Variablen der Clusteranalyse übergewichtet.

Eine mögliche *Übergewichtung des Zeitaspekts* kann auch die Einordnung der *drei Probanden aus Profil 2 in den Cluster 3* erklären. Die Probanden Nr. 8, 6 und 19 weisen ähnlich kurze Trainings- und Gesamtspielzeiten auf wie einige der Probanden aus Gruppe 3 (Nr. 2, 9, 20, 21, 25, 30, 34). Werden alle Probanden in Bezug auf die Trainingszeit in eine gemeinsame Rangreihe gebracht, so befinden sich die genannten sieben Probanden aus Profil 3 zusammen mit den drei genannten Probanden aus Profil 2 auf den Rangplätzen 1 bis 11 (ein Platz wird von einem Probanden aus Profil 1 eingenommen). Auch hier hat demnach das Merkmal *Zeit* (Trainings-, Gesamtspielzeit) einen stärkeren Einfluss auf die Gruppenbildung als das Merkmal *Kurvenverlauf* (Ereignispunkte X, M2, M3, M4 und Ende).

Beim vierten Probanden (Nr. 24), der bei der Ward-Methode ebenfalls in den Cluster 3 (statt 2) eingeordnet wird, handelt es sich auch bei der inhaltlich orientierten Klassifizierung um einen *Grenzfall*. Beide Kriterien für die Einordnung ins Profil 2

(Höhe des Endwerts, Anstieg während Testphase) werden von diesem Probanden nur knapp erreicht.

Die Anwendung der *k-means-Methode* auf die Ergebnisse der Clusteranalyse ergibt im Grossen und Ganzen keine neuen Erkenntnisse. Wiederum bilden die vier Probanden mit den langen Trainingszeiten (Profil 3.3) einen eigenen Cluster. Einige Probanden aus den übrigen drei Clustern werden umgeteilt, was inhaltlich jedoch weniger plausibel erscheint. Das Gleiche gilt für die Optimierung der Drei-Clusterlösung bei der Variante ohne Ausreisser ($n = 31$).

Die *Klassifizierung anhand der Euklidischen Distanzen* (MDS und Clusteranalyse) bestätigt die inhaltlich orientierte Klassifizierung noch deutlicher als die Clusteranalyse über die Ereignispunkte. Alle Profil-1-Probanden grenzen sich deutlich von den übrigen Probanden ab, und die Mehrheit der Profil-2- sowie Profil-3-Probanden wird ebenfalls als zusammengehörig abgebildet. Wie bei der Clusteranalyse über die Ereignispunkte werden einige Profil-2-Probanden zusammen mit Cluster-3-Probanden abgebildet. Abbildung 9 zeigt beispielhaft das Ergebnis der MDS für die „Fixed-Step“ Variante. In die abgebildete Konfiguration sind das Ergebnis der entsprechenden Clusteranalyse sowie die Gruppenbildung aufgrund der inhaltlich orientierten Klassifizierung eingetragen.

Die Einordnung der 35 Probanden in den zweidimensionalen Darstellungsraum gelingt bei allen drei Varianten zur Berechnung der Euklidischen Distanzen sehr gut: Die MDS wird jeweils nach vier bis fünf Iterationen abgebrochen, da sich das Gütekriterium (S-stress nach Young) in der letzten Iteration um weniger als .0001 verändert. Der Ausprägungsgrad des Gütekriteriums der resultierenden Konfiguration (Kruskal's Stress) liegt zwischen .03 und .08, der Ausprägungsgrad der quadrierten Korrelation zwischen den Ähnlichkeitsmassen und den Distanzen (RSQ) liegt zwischen .980 und .999. Das bedeutet, dass praktisch die gesamte Varianz in den Ähnlichkeitsmassen (d.h. den Euklidischen Distanzen zwischen den Kurven) durch die Varianz in den Distanzen (MDS) repräsentiert wird.

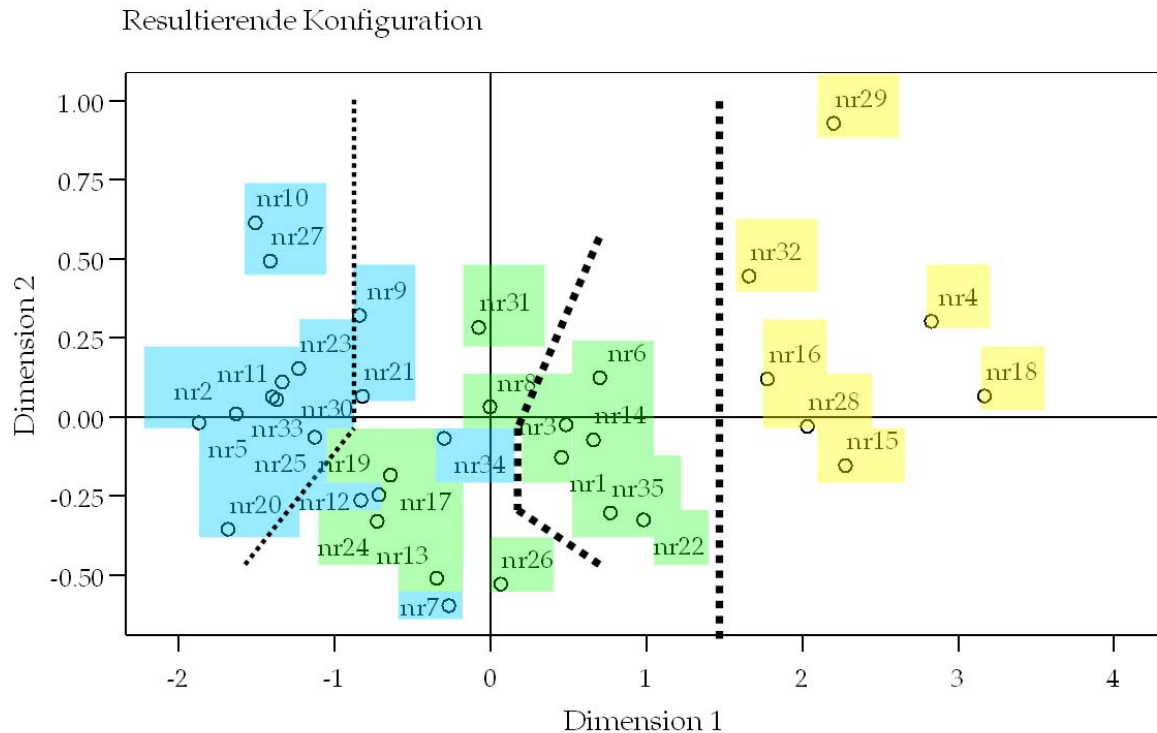


Abbildung 9: Ergebnis der MDS für das Informationseröffnende Handeln (Berechnung der Euklidischen Distanzen über Variante „Fixed Step“). Die eingezeichneten gestrichelten Linien symbolisieren das Ergebnis der entsprechenden Clusteranalyse (Drei- bis Vier-Clusterlösung). Die Einfärbungen kennzeichnen die Gruppeneinteilung aufgrund der inhaltlich orientierten Klassifizierung, wobei gelb Profil 1, grün Profil 2 und blau Profil 3 markiert (n = 35).

Die Ergebnisse der entsprechenden *Clusteranalysen* empfehlen eine Lösung mit drei bis vier Clustern, wobei bei den Varianten „Fixed Step“ und „Fixed Measures“ alle Probanden aus Profil 1 einen eigenen Cluster bilden (vgl. Abbildung 9). Probanden des Profils 2 werden aufgeteilt, sechs bilden einen eigenen Cluster und sieben (Nr. 8, 13, 17, 19, 24, 26, 31) werden dem dritten Cluster zugeordnet, genauer zu fünf Probanden aus Profil 3. Die anderen zehn Profil-3-Probanden bilden den vierten Cluster, der bei der Drei-Clusterlösung mit dem Cluster 3 fusioniert wird. Bei allen Berechnungsvarianten bilden die vier Probanden mit den langen Verläufen (Nr. 5, 7, 23 und 27) – im Unterschied zur Clusteranalyse über die Ereignispunkte – keinen eigenen Cluster, sondern werden dem Cluster 3 beziehungsweise 4 zugeteilt.

Wie bei der Clusteranalyse über die Ereignispunkte können die Abweichungen von der inhaltlich orientierten Klassifizierung erklärt werden. Bei vier der Probanden aus Profil 3, die zusammen mit Probanden aus Profil 2 einen Cluster bilden (Nr. 9, 12, 21,

34), handelt es sich wiederum um *Grenzfälle*. Alle haben einen Endwert, der für eine Einteilung in Profil 2 sprechen würde (> 45), zeigen während der Testphase jedoch eine zu geringe Steigung (< 30). Proband Nr. 7 erfüllt beide Kriterien für Profil 2, „rutscht“ jedoch aufgrund der langen Trainingszeit ins Profil 3.

4.3.2.3 Lösung und Interpretation für das Adäquate Handeln

Bei der *Klassifizierung anhand der Ereignispunkte* empfehlen alle Clusteranalyse-Varianten eine Lösung mit drei Clustern, wobei jeweils die Profile 2 und 3 zu einem Cluster vereinigt sind. Eine Vier-Clusterlösung ist aber ebenfalls vertretbar. Bei allen gerechneten Clusteranalyse-Varianten bilden die Probanden des Profils 1 einen eigenen Cluster. Auch die Mehrheit der Probanden der Profile 2, 3 und 4 bildet je einen eigenen Cluster. Je nach Berechnungsvariante (verwendeter Algorithmus, Ausschluss von Ausreißern) weicht das Ergebnis der Clusteranalysen insgesamt für *sechs bis acht Probanden* von der inhaltlich orientierten Klassifizierung ab. Diese Abweichungen sind bei einer genaueren Betrachtung wiederum einfach zu erklären. Im Folgenden werden beispielhaft die sieben Abweichungen erläutert, die sich bei der Ward-Methode und der Berücksichtigung aller Probanden ergeben (vgl. Abbildung 10).

Die *zwei Probanden aus dem Profil 4*, die dem Cluster 3 zugeordnet werden (Nr. 3, 31), haben beide einen *degressiv ansteigenden Verlauf* mit einem relativ hohen Trainingsendwert von mehr als 20 Punkten und einem um diesen Wert schwankenden Verlauf während der Testphase. Ohne eine Bedeutungszuweisung, das heisst ohne die separate Betrachtung der Trainings- und Testphase, würden beide Verläufe auch bei der inhaltlich orientierten Klassifizierung ins Profil 3 passen (Endwert > 20). Bei der Clusteranalyse erfolgt die Gruppenbildung über ähnliche Ausprägungsgrade zu den Zeitpunkten X, M2, M3, M4 und Ende, wobei die beiden Profil-4-Probanden in vier der fünf Variablen ähnliche Ausprägungsgrade wie die Cluster-3-Probanden haben. Der (für den Cluster 3 zu) hohe *Trainingsendwert* der beiden Probanden wird *untergewichtet*.

Dendrogram using Ward Method

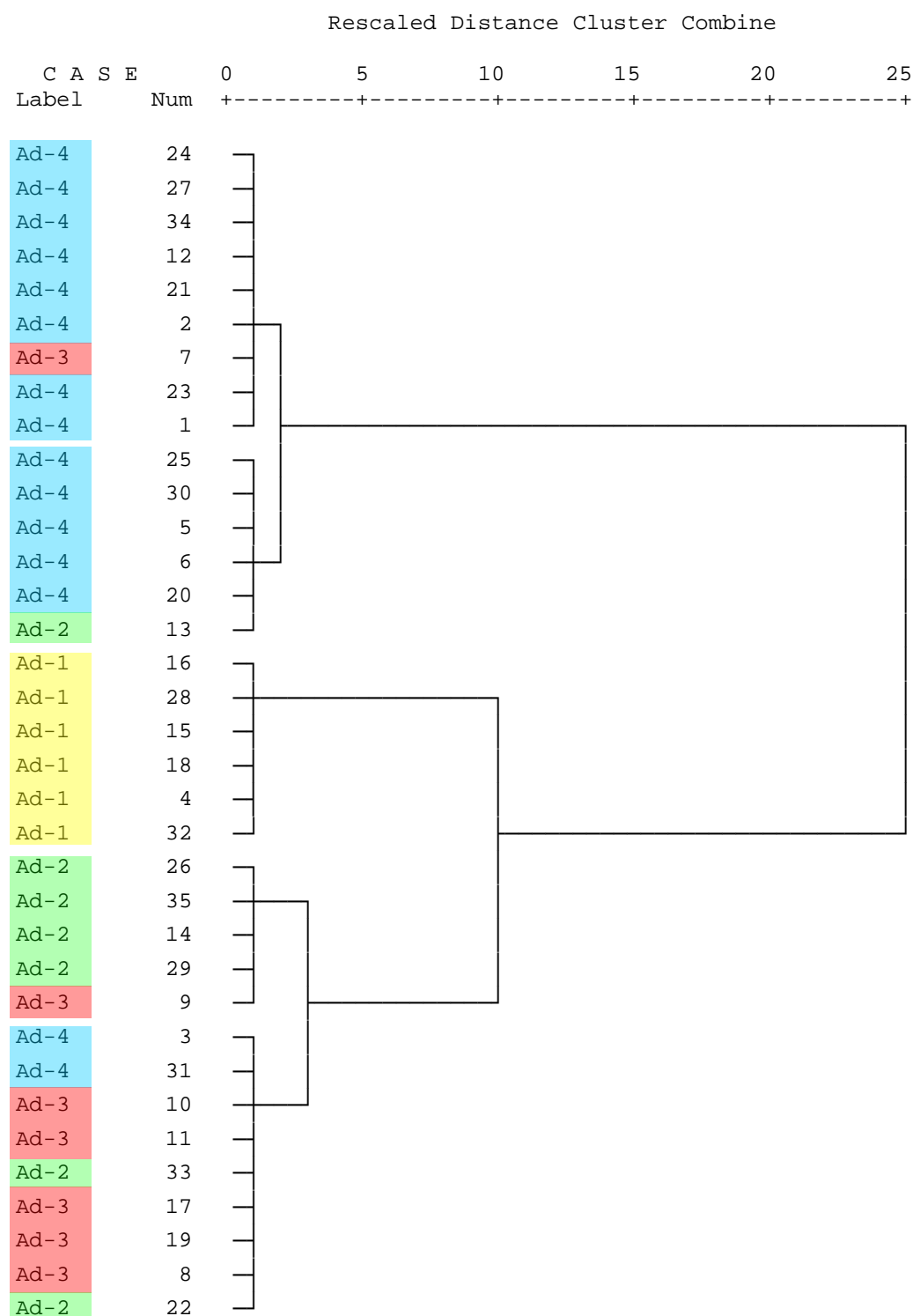


Abbildung 10: Dendrogramm der Clusteranalyse über die Ereignispunkte für das Adäquate Handeln. Die Probanden sind durch die Gruppeneinteilung entsprechend der inhaltlich orientierten Klassifizierung gekennzeichnet, wobei gelb Profil 1, grün Profil 2, rot Profil 3 und blau Profil 4 markiert (n = 35).

Ähnliches gilt für die *drei Probanden aus Profil 2*, die entweder dem *Cluster 3* (Nr. 33, 22) oder sogar dem *Cluster 4* (Nr. 13) zugeordnet werden. Es handelt sich um die drei Probanden mit *Startschwierigkeiten im Training*, das heisst flachen (Nr. 33, 22) oder abfallenden (Nr. 13) Verläufen während der Trainingsphase und steil ansteigenden Verläufen während der Testphase. Die Werte während der Testphase sind aufgrund der schlechten Startphase niedriger als bei den anderen Probanden aus Profil 2, was die Zuordnung zu den Clustern 3 und 4 erklärt. Die hohe Ähnlichkeit in vier von fünf Variablen wird wiederum übergewichtet (bzw. die Unähnlichkeit im Zeitpunkt X untergewichtet). Während die inhaltlich orientierte Klassifizierung den Gesamtverlauf, und vor allem die unterschiedlichen Verläufe während Trainings- und Testphase mitberücksichtigt, wird bei der Clusteranalyse der *Verlauf während der Testphase übergewichtet*. Auch durch die Berücksichtigung weiterer Variablen (wie beispielsweise der Info-Zuwachs während der Testphase oder die Steigung in den beiden Spielphasen) lässt sich die inhaltlich orientierte Klassifizierung durch die Clusteranalyse nicht besser abbilden. Auf die Berichte dieser zusätzlichen Analysen wird an dieser Stelle deshalb verzichtet.

Die *zwei Probanden aus Profil 3*, die entweder dem *Cluster 2* (Nr. 9) oder dem *Cluster 4* (Nr. 7) zugeordnet werden, gehören auch in der inhaltlich orientierten Klassifizierung zu *Grenzfällen*. Proband Nr. 9 „verpasst“ die Einordnung ins Profil 2 relativ knapp um 3 Punkte (47 statt 50 Punkte). Der Verlauf ist ab dem Zeitpunkt M2 tendenziell abflachend, wobei die Werte zu den Zeitpunkten X, M2, und M3 annähernd so hoch sind wie bei einigen Probanden aus Profil 2. Dies führt bei der Clusteranalyse wiederum zu einer Übergewichtung dieser Gemeinsamkeit.

Proband Nr. 7 zeigt bis zum Zeitpunkt M3 einen wechselnden, flachen Verlauf (der für eine Einordnung ins Profil 4 spricht), danach steigt die Kurve steil um mehr als 25 Punkte an, weshalb er bei der inhaltlich orientierten Klassifizierung ins Profil 3 „rutscht“. Bei der Clusteranalyse wird er aufgrund der eher tiefen Werte in X, M2, M3 und M4 dem Profil 4 zugeordnet. Allerdings wird er – als einziger Proband überhaupt – durch das anschliessende *k-means-Verfahren* zur Optimierung der gefundenen Clusterlösung umgeteilt, und zwar entsprechend der inhaltlich orientierten Klassifizierung in den Cluster 3.

Die Klassifizierung anhand der Euklidischen Distanzen (MDS und Clusteranalyse) ergibt praktisch die gleichen Ergebnisse wie die Clusteranalyse über die Ereignispunkte. Es zeigen sich nur zwei Unterschiede: Proband Nr. 17 wird bei allen drei Berechnungsvarianten in den Cluster 4 (statt 3) eingeteilt, und bei der Variante „Fixed Step“ wird Proband Nr. 9 – entsprechend der inhaltlich orientierten Klassifizierung – dem Cluster 3 zugeordnet. Aufgrund der praktisch identischen Lösung wird auf eine detaillierte Beschreibung der Gruppeneinteilung (bzw. der Abweichung von der inhaltlich orientierten Klassifizierung) verzichtet.

Beispielhaft zeigt Abbildung 11 das Ergebnis der MDS für die „Fixed-Step“ Variante. In die abgebildete Konfiguration sind wiederum das Ergebnis der entsprechenden Clusteranalyse sowie die Gruppenbildung aufgrund der inhaltlich orientierten Klassifizierung eingetragen.

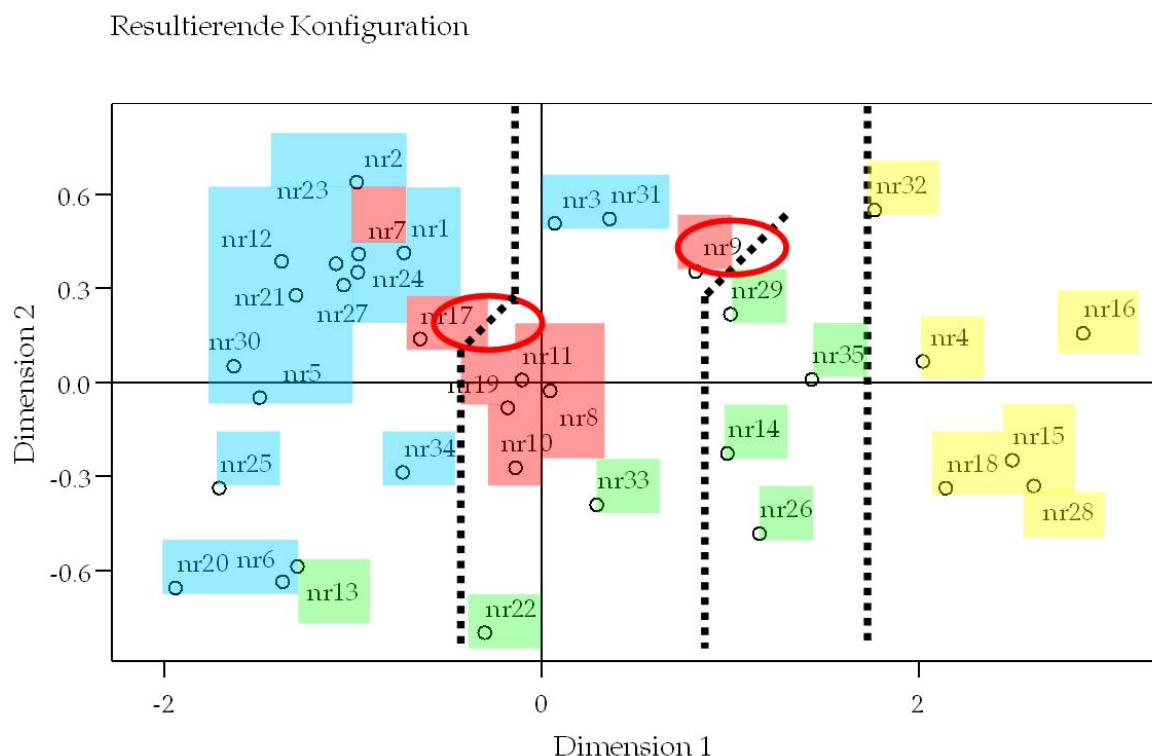


Abbildung 11: Ergebnis der MDS für das Adäquate Handeln (Berechnung der Euklidischen Distanzen über Variante „Fixed Step“). Die eingezeichneten gestrichelten Linien symbolisieren das Ergebnis der entsprechenden Clusteranalyse (Vier-Clusterlösung). Die Einfärbungen kennzeichnen die Gruppeneinteilung aufgrund der inhaltlich orientierten Klassifizierung, wobei gelb Profil 1, grün Profil 2, rot Profil 3 und blau Profil 4 markiert. Unterschiede zur Clusteranalyse über die Ereignispunkte sind mit roten Kreisen markiert ($n = 35$).

Wie beim Informationseröffnenden Handeln gelingt die Einordnung der 35 Probanden in den zweidimensionalen Darstellungsraum bei allen Berechnungsvarianten äusserst gut. Die MDS bricht nach drei bis fünf Iterationsschritten ab, wobei die Gütekriterien praktisch die gleichen Ausprägungsgrade haben wie beim Informationseröffnenden Handeln (.03 bis .08 für Kruskal's Stress und .978 bis .998 für RSQ). Auch hier kann die Varianz in den Ähnlichkeitsmassen fast zu 100% durch die Varianz in den Distanzen repräsentiert werden.

4.3.2.4 Zusammenfassung und Diskussion

Die aufgrund inhaltlicher Überlegungen definierten Prozessprofile der beiden Konstrukte lassen sich durch die formalen Klassifizierungsmethoden recht gut bestätigen. Vor allem die *Klassifizierung der jeweils besten Probanden* (Prozessprofile 1) gelingt bei beiden Konstrukten und bei *unterschiedlichen Operationalisierungen der Ähnlichkeit* (Ereignispunkte, Euklidische Distanzen mit verschiedenen Berechnungsvarianten) sowie *unterschiedlichen Klassifizierungsmethoden* (Clusteranalyse-Methoden Ward, Within-groups-/Between-groups-linkage, Single-linkage, k-means; Multidimensionale Skalierung) sehr gut. Die Ergebnisse der verschiedenen Zugänge werden im Folgenden kurz zusammengefasst und bewertet.

Beim *Informationseröffnenden Handeln* ergibt sich bei der Clusteranalyse über die *Ereignispunkte* eine Verzerrung infolge Übergewichtung des Zeitaspekts, dadurch dass sowohl die Gesamtspielzeit wie auch die Trainingszeit als zusätzliche Variablen in die Clusteranalyse einbezogen wurden. Werden Probanden mit sehr langen Verläufen (Ausreisser gemäss Single-linkage-Methode) von der Analyse ausgeschlossen, zeigt sich *praktisch die gleiche Drei-Clusterlösung wie bei der inhaltlich orientierten Klassifizierung*. Der einzige Unterschied besteht in der Zuordnung von (je nach Methode) *drei bis vier Probanden aus Profil 2 zum Cluster 3*. Diese „Fehlzuordnungen“ lassen sich erklären: a) als Grenzfall bei der inhaltlich orientierten Klassifizierung (Nr. 24) und b) wiederum durch die Übergewichtung des Zeitaspekts (Nr. 8, 6, 19 mit gleich kurzen Trainingszeiten wie Probanden aus dem Cluster 3).

Bei der Klassifizierung anhand der *Euklidischen Distanzen* (MDS, Clusteranalysen) lassen sich die drei Profile auch ohne Ausschluss von Ausreißern deutlich abbilden. Die Clusteranalysen ergeben zwar eine Vier-Clusterlösung, eine Drei-Clusterlösung (mit Fusion der Cluster 3 und 4) ist aber ebenfalls vertretbar. Die Abweichung von der inhaltlich orientierten Klassifizierung besteht in der *Aufteilung der Profil-2-Probanden* zu zwei Clustern, wobei sieben der 13 Profil-2-Probanden (Nr. 8, 13, 17, 19, 24, 26, 31) zusammen mit fünf Profil-3-Probanden (Nr. 7, 9, 12, 21, 34) einen Cluster bilden.

Bemerkenswert ist, dass sowohl bei der Klassifizierung über die Ereignispunkte wie auch bei der Klassifizierung über die Euklidischen Distanzen *Profil 1 eindeutig als alleiniger Cluster erkannt* wird. Auch bei der MDS grenzen sich die sieben Profil-1-Probanden klar von den anderen, schlechteren 28 Probanden ab. Bei beiden Zugängen wird ein Teil der Profil-2-Probanden der Gruppe 3 zugeordnet, jedoch handelt es sich je nach Zugang und Clusteranalyse-Methode um andere Probanden.

Beim *Adäquaten Handeln* zeigen alle Berechnungsvarianten (Ereignispunkte, Euklidische Distanzen in drei Varianten, Clusteranalysen, MDS) mit Ausnahme von zwei Probanden (Nr. 17, 9) *die gleiche Lösung*. Im Vergleich zur inhaltlich orientierten Klassifizierung ergeben sich sieben (Ward-Methode) beziehungsweise sechs (k-means-Verfahren) Abweichungen, die *je zwei bis drei Probanden der Cluster 2, 3 und 4 betreffen* (Nr. 3, 7, 9, 13, 17, 22, 31, 33).

Wiederum werden *Profil-1-Probanden eindeutig als alleiniger Cluster abgebildet*, und die Abweichungen in den anderen Profilen können durch Übergewichtung bestimmter Aspekte bei der Clusteranalyse erklärt werden. Letzteres betrifft hier die Übergewichtung des Verlaufs während der Testphase beziehungsweise die Untergewichtung des Verlaufs während der Trainingsphase oder frühen Testphase. Bei der inhaltlich orientierten Klassifizierung wird zwar ebenfalls die Testphase stärker gewichtet, wobei jedoch vor allem der *am Trainingsendwert rekali-brierte Verlauf* bewertet wird.

Für die Beurteilung der Ähnlichkeit der Zeitreihen wurden sehr unterschiedliche Methoden angewendet. Während die Clusteranalyse über die Ereignispunkte –



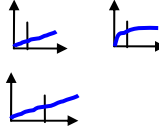









analog zur inhaltlich orientierten Klassifizierung – die Ähnlichkeit hauptsächlich in Bezug auf den *Spielverlauf* beurteilt (Ähnlichkeit zu bestimmten, ausgewählten Ereignissen), beurteilt die Clusteranalyse beziehungsweise die MDS über die Euklidischen Distanzen die *absolute Ähnlichkeit* der Zeitreihen unabhängig vom Spielgeschehen (ähnlich = möglichst deckungsgleich). Für die Berechnung der Euklidischen Distanzen (d.h. der Abstände zwischen den Zeitreihen) kamen dabei drei Berechnungsvarianten zur Anwendung: Zwei Varianten berücksichtigen in unterschiedlicher Weise die *absolute Spielzeit*, das heisst die unterschiedliche Länge der Zeitreihen, eine Variante berücksichtigt nur die *Verlaufsform*, unabhängig von deren Länge.

Trotz der unterschiedlichen Ähnlichkeitsformalisierungen und Methoden zur Ähnlichkeitsbeurteilung ergaben sich im Grossen und Ganzen sehr ähnliche Lösungen, die zudem – mit einigen Ausnahmen – mit der inhaltlich orientierten Klassifizierung übereinstimmen. Die „Fehlzuordnungen“ lassen sich mehrheitlich durch Methodenaspekte erklären (insbesondere durch Übergewichtung bestimmter Merkmale infolge hoher Interkorrelationen von Variablen) und betreffen zudem (vor allem beim Informationseröffnenden Handeln) bei den verschiedenen methodischen Zugängen nicht immer die gleichen Probanden. Deshalb wird die inhaltlich orientierte Klassifizierung beibehalten.

4.3.3 Beurteilungsschema und Problemlösetypen: Beschreibung und Diskussion

Die Kombinationen der definierten und durch formale Methoden bestätigten Prozessprofile bilden das Beurteilungsschema. Die Bedeutung der Problemlösetypen ergibt sich aus den Interpretationen der einzelnen Prozessprofile. Da diese in Kapitel 4.3.1 bereits erläutert worden sind, werden die Problemlösetypen nicht im Detail beschrieben, sondern lediglich einzelne Aspekte der Profil-Kombinationen hervorgehoben. Tabelle 5 zeigt, wie sich die Probanden der Teilstichprobe 1 auf die verschiedenen Problemlösetypen verteilen.

Tabelle 5: Verteilung der Problemlösetypen in der Teilstichprobe 1.

		Prozessprofile [†] für das Informationseröffnende Handeln (I)			Total
		1	2	3	
					
Prozessprofile [†] für das Adäquate Handeln (A)	1 	6 (17%)	0	0	6 (17%)
	2  	1 (3%)	5 (14%)	1 (3%)	7 (20%)
	3  	0	3 (9%)	4	7 (20%)
	4  	0	5 (14%)	10 (29%)	15 (43%)
	 				
Total		7 (20%)	13 (37%)	15 (43%)	35 (100%)

† *Bemerkung:* Die Prozessprofile beinhalten teilweise mehrere typische Verlaufsformen, die hier exemplarisch abgebildet sind (für genauere Angaben vgl. Kapitel 4.3.1).

Wie erwartet kommen bestimmte Kombinationen nicht vor, nämlich die Kombinationen aus dem jeweils besten und dem schlechtesten (bzw. dem zweitschlechtesten bei Ad) Profil (Typen I1-A4, I1-A3, A1-I3). Die Kombinationen der *Prozessprofile 1 und 2 der beiden Konstrukte* beschreiben die *guten bis sehr guten Problemlöser*. Sechs Probanden gehören zu den sehr guten und weitere sechs zu den guten Problemlösern. Ihr Vorgehen ist optimal und effizient, sie verschaffen sich mit fast jedem Handlungsschritt und durch verschiedenste Aktionen neue Informationen und wissen diese auch anzuwenden. Für den Problemlösetyp I2-A2 kann sich dieses optimale Problemlöseverhalten auch nur auf die Testphase beziehen, während das Handeln während der Trainingsphase wenig bis nicht sinnvoll sein kann (Ad-Profil-typ 2.2 mit Startschwierigkeiten, vgl. Kapitel 4.3.1.2).

Als Führungskraft ungeeignet ist der *Problemlösetyp I3-A4*. Diese Probanden verschaffen sich kaum (oder nur sehr ineffizient oder nur während dem Training) problemrelevante Informationen und können ihr Vorgehen nicht der aktuellen Aufgabenstellung anpassen. Such- und Prüfinstrumente werden nicht korrekt eingesetzt (z.B. Suche im Wasser statt auf Meeresgrund), so dass sich die Informationsaufnahme vor allem auf das Lesen beschränkt. Das inadäquate Verhalten dieser Probanden kann alle Teilsysteme betreffen oder nur einzelne (wie die Kommunikation oder die Steuerung), und von Anfang an vorhanden sein oder erst mit zunehmend schwierig werdender Aufgabe (während der Testphase). Letzteres gilt auch für den *Problemlösetypen I2-A4*, wobei es diesen Probanden aber immerhin gelingt, sich immer wieder (vor allem aufgrund eines erfolgreicherer Einsatzes der Such- und Prüfinstrumente) neue Informationen zu beschaffen. Im Unterschied zum *Problemlösetyp I2-A3* können sie diese jedoch weniger gut umsetzen.

Die verschiedenen Problemlösetypen unterscheiden sich – zusammengefasst und vereinfacht interpretiert – darin, wie effizient und konsistent für die Problemlösung relevante Informationen gesucht werden, und inwiefern die (potentiell) aufgenommenen Informationen umgesetzt und genutzt werden können. Wirth und Funke (2005) sprechen in diesem Zusammenhang vom Identifizieren und Integrieren von Informationen und von Wissenserwerb und Wissensanwendung (vgl. Kapitel 2.1.2). Erfolgreiche und nicht erfolgreiche Problemlöser unterscheiden sich für die Autoren darin, dass...

...es wenig erfolgreichen Problemlösern offensichtlich schwer fällt, einmal entdeckte oder generierte Informationen für spätere (Anwendungs-)Gelegenheiten festzuhalten. Der Aufwand, der für das Identifizieren betrieben wird, verpufft, weil einmal identifizierte Informationen wieder in Vergessenheit geraten. In späteren Situationen steht der Problemlöser damit wieder vor genau denselben Problemen, er bleibt sozusagen ein ewiger Problemlöser, während erfolgreiche Problemlöser im Umgang mit dem System dazulernen und die Steuerung des Systems immer mehr zu Routine wird. (S. 67)

Wenig erfolgreiche Problemlöser identifizieren und integrieren Informationen ausserdem nur dann, wenn sie ihnen vom System dargeboten werden, das heisst sie handeln *wenig selbstreguliert*, während sich erfolgreiche Problemlöser von Anfang an auf die Identifikation und Integration neuer Information konzentrieren (Wirth, 2004).

Dies bestätigen auch die eigenen Daten: Info-Profil-3-Probanden (genauer: jene mit flach ansteigendem Verlauf, Profil 3.1, vgl. Kapitel 4.3.1.1) beschränken ihre Informationsaufnahme vorwiegend auf das Lesen eingehender Meldungen. Die Ankunft solcher Meldungen ist jeweils mit einem während längerer Zeit blinkenden Warnsignal versehen, das heisst der Proband wird explizit aufgefordert, die Meldung zu lesen. Zusätzliche Informationen, die nicht unmittelbar zugänglich sind (Datenbank Nauti-Info oder Hilfesystem), werden von Info-Profil-3-Probanden weniger erkannt, und der Einsatz von Sonar- und Videogerät bleibt entweder aufgrund falscher Anwendung informationsleer oder erfolgt zu spärlich. Demgegenüber suchen Info-Profil-1-Probanden sowohl während der Trainings- wie auch der Testphase aktiv nach neuen Informationen, indem sie das Hilfesystem sowie die Datenbank Nauti-Info gezielt nach neuen Informationen durchsuchen und kontinuierlich den Meeresboden nach Metallfunden absuchen. Fast jeder Handlungsschritt ist darauf ausgerichtet, problemrelevante Information zu verschaffen. Probanden der Problemlösetypen I1-A1/A2 können diese Information auch behalten und korrekt umsetzen, so dass sowohl Info- wie auch Ad-Kurve (während der Testphase) steil ansteigend verlaufen und wenig auseinander klaffen.

Andere Erklärungen für die mangelnde bis schlechte Informationsaufnahme liefern Untersuchungen zu den Fehlern beim Umgang mit komplexen Situationen (Schaub, 2006). Demnach wird häufig nur die Information beschafft, die in das eigene Modell passt, während Widersprüche und Ähnliches nicht wahrgenommen werden („*Kanalisierung der Informationssammlung*“), oder dem eigenen Modell widersprechende Informationen werden ausgeblendet, auch wenn das Modell noch so einfach ist („*Hypothesengerechte Informationsauswahl*“).

Ein solches Verhalten ist im neuen Auswertungsmodell daran zu erkennen, dass der Verlauf für das Adäquate Handeln stark und unter Umständen abrupt abfällt, während der Verlauf für das Informationseröffnende Handeln – durch das Lesen eingehender Meldungen oder anderer relevanter Informationen – noch minimal ansteigen kann. Ein sehr illustratives Beispiel hierfür ist *Proband Nr. 2 der Teilstichprobe 1*, der kurz nach Erhalt der dritten Notfallmeldung (M3) einen Fund im Gebiet R/8 ortet und sich anschliessend darauf versteift, dass es sich dabei um das Flugzeug

handeln muss, auch wenn neuere Informationen (Notfallmeldung M4 mit dem Hinweis auf einen Fund im Gebiet S/7) dieser Hypothese widersprechen. Ganz ähnlich verhält sich *Proband Nr. 5 der Teilstichprobe 2*, der nach ein paar wenigen Sonar- und Videoaufnahmen auf Grund im Gebiet S/7 an der Hypothese festhält, das Flugzeug gefunden zu haben. Widersprechende Informationen (Notfallmeldungen M3 und M4 mit Hinweis auf einen Absturz im Gebiet R/9) werden ausgeblendet und der Proband verbringt die restlichen gut 20 Spielminuten damit, am genau gleichen (!) Ort Sonar- und Videoaufnahmen zu machen und Helfer anzufordern.

Eine vollständige Beschreibung und Interpretation der Problemlösetypen kann und soll an dieser Stelle nicht erfolgen. Für einen Einsatz der neuen Auswertung in der diagnostischen Praxis braucht es natürlich detaillierte und mehr praxisbezogene Angaben zu den Problemlösetypen. Bevor diese ausgearbeitet werden, muss jedoch die neue Auswertung zumindest ansatzweise auf ihre Gültigkeit überprüft werden. Dies ist Bestandteil der folgenden Abschnitte.

4.4 Überprüfung des Auswertungsmodells

Als erstes wird überprüft, ob die neue Auswertung das Gleiche misst wie die bestehende, szenariospezifische Auswertung (genauer: wie die darin enthaltenen 13 Leistungskennwerte; Fragestellung 4). Anschliessend wird die neue Auswertung dem Urteil eines Experten gegenübergestellt, welcher die Ergebnisse der szenariospezifischen Auswertung interpretiert und (zu einem ebenfalls szenario-unspezifischen Urteil) integriert hat (Kapitel 4.4.2; Fragestellung 5). Das letzte Kapitel 4.4.3 überprüft die Stichprobenabhängigkeit der durch das neue Auswertungsmodell definierten Interpretations- und Klassifikationsregeln (Fragestellung 6).

Im Unterschied zu den vorangegangenen Kapiteln werden für alle Fragestellungen nur die Ergebnisse berichtet. Deren Diskussion folgt erst in Kapitel 5.

4.4.1 Vergleich mit der szenariospezifischen Auswertung

Die Verteilungen der 13 Leistungskennwerte in der Teilstichprobe 1 zeigen, dass drei Kennwerte die Probandenleistung nur wenig differenzieren. Bei den Kennwerten „Bearbeitungszeit für abgerufene Informationen effizient nutzen (p_{ow_i})“, „Informationen effizient aus- und weitergeben ($p_{m_p_nv}$)“ und „Ankommende Informationen vollständig abrufen (p_{iseit})“ liegt der Median über 90 und die Interquartilweite ist deutlich kleiner als bei den anderen Kennwerten (vgl. Abbildung 12). Der Kennwert p_{ow_i} ist zudem der einzige Kennwert mit fehlenden Werten: Sieben Probanden haben während der Problembearbeitung nicht in der Datenbank Nauti-Info gelesen.

Die drei in der Teilstichprobe 1 wenig differenzierenden Kennwerte werden für den Vergleich mit dem neuen Auswertungsmodell deshalb ausgeschlossen.

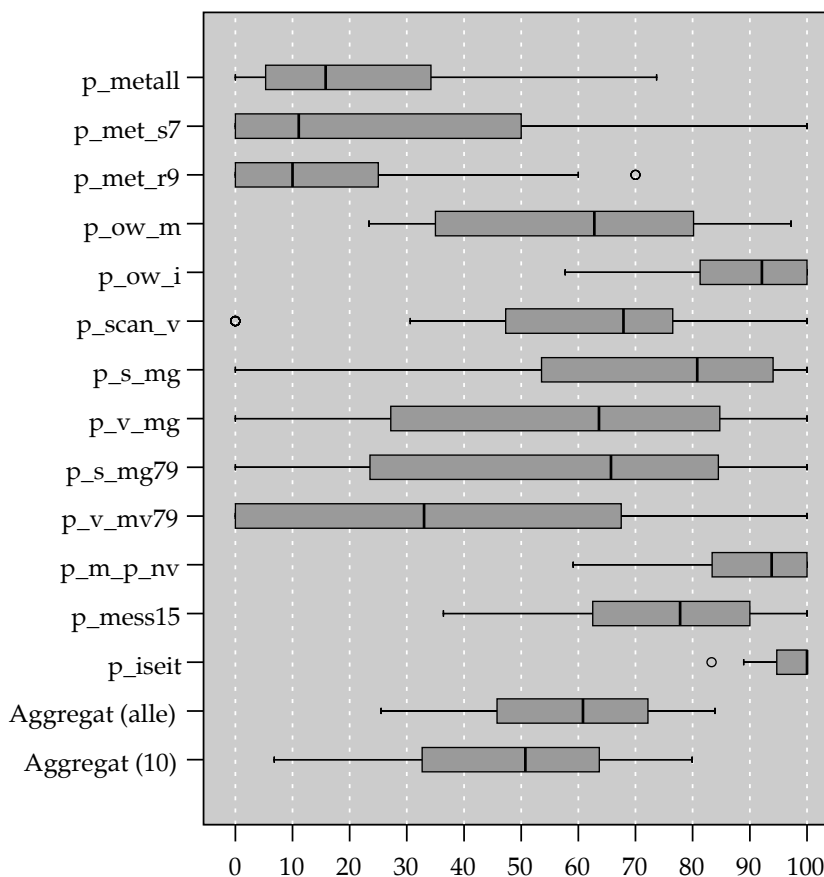


Abbildung 12: Verteilung der 13 Leistungskennwerte in der Teilstichprobe 1 ($n = 35$, für p_{ow_i} und Aggregat über alle Kennwerte ist $n = 28$).

Aufgrund der multivariaten Reliabilitäts- und Validitätstheorie (Wittmann, 1987; Wittmann & Matt, 1986) ist zu erwarten, dass der Zusammenhang zwischen Leistungskennwerten und dem neuen Auswertungsmodell höher ist, wenn erstere aggregiert werden (gleiches Generalitätsniveau; vgl. Kapitel 2.2.3.3). Deshalb werden zusätzlich zwei aggregierte Masse gebildet, einmal die Summe aus allen 13 Leistungskennwerten ($n = 28$ wegen fehlender Werte in p_{ow_i}) und einmal die Summe aus den zehn gut differenzierenden Kennwerten (ohne p_{ow_i} , $p_{m_p_nv}$, p_{iseit}). Die aggregierten Werte werden wieder in die Skala der Leistungskennwerte abgebildet (0-100). Angesichts der unterschiedlichen Spannweite und Streuung der einzelnen Leistungskennwerte (vgl. Abbildung 12), wäre eine Aufsummierung der z-transformierten Werte angezeigt. Da die Werte aber weder in der Teilstichprobe 1 noch in der Gesamtstichprobe normalverteilt sind, wird darauf verzichtet.

Der Zusammenhang zwischen den *Prozessmassen* (d.h. den Endwerten für das Informationseröffnende und Adäquate Handeln) und den Leistungskennwerten ist wie erwartet positiv und *für neun der zehn geprüften Kennwerte signifikant* (Hypothese 4.1). Die Korrelationen liegen zwischen .43 und .74 für das Informationseröffnende Handeln und zwischen .31 und .74 für das Adäquate Handeln (vgl. Tabelle 6). Damit liegen nach der Effektstärkenklassifikation von Cohen (1992, S. 157) mittlere und mehrheitlich sogar starke Effekte vor. Bei beiden Konstrukten können jeweils für den Kennwert „Suchaktivitäten in sinnvoller Tiefe (p_{s_mg})“ die höchsten und für den Kennwert „Ankommende Informationen effizient aufnehmen (p_{mess15})“ die geringsten Korrelationen beobachtet werden, und bei beiden Konstrukten lässt sich für den Kennwert „Bearbeitungszeit für eingegangene Informationen effizient nutzen (p_{ow_m})“ kein signifikanter Zusammenhang nachweisen.

Wie erwartet fallen die Korrelationen für die *aggregierten Masse* höher aus: Für das Aggregat über alle Leistungskennwerte ergeben sich Korrelationen von über .84. Werden nur die zehn Kennwerte berücksichtigt, die die Probandenleistung genügend differenzieren, so liegen die Korrelationen etwas tiefer, sie sind aber immer noch höher als die Korrelationen mit den einzelnen Leistungskennwerten (mit Ausnahme von p_{s_mg} , wo ebenfalls Korrelationen von .74 vorliegen).

Tabelle 6: Zusammenhang zwischen den szenariospezifischen Leistungskennwerten und Informationseröffnendem sowie Adäquatem Handeln (n = 35).

	Informationseröffnendes Handeln		Adäquates Handeln	
	r_s^\dagger	$J-T^\ddagger$	r_s^\dagger	$J-T^\ddagger$
p_metall	.67***	-4.75***	.53**	-3.34**
p_met_s7	.40**	-3.14*	.34*	-2.06*
p_met_r9	.63***	-3.36**	.46***	-2.76*
p_ow_m	.13 (p = .23)	-.77 (p = .44)	.27 (p = .06)	-1.56 (p = .12)
p_scan_v	.50**	-2.45*	.43**	-3.02*
p_s_mg	.74***	-4.27***	.74***	-4.19***
p_v_mg	.56***	-3.12*	.68***	-3.51***
p_s_mg79	.59***	-3.48**	.60***	-3.28**
p_v_mv79	.65***	-3.28**	.62***	-3.20**
p_mess15	.43**	-3.25**	.31*	-2.16*
Aggregiertes Mass (alle Kennwerte, n = 28)	.85**	-5.0***	.87**	-4.87***
Aggregiertes Mass (10 Kennwerte)	.74**	-4.62***	.73**	-4.49***

† Prüfung des Zusammenhangs zwischen dem Prozessmass (Endwert) und den Leistungskennwerten anhand der Korrelation nach Spearman bei einseitiger Hypothesenprüfung.

‡ Prüfung des Zusammenhangs zwischen den Prozessprofilen und den Leistungskennwerten anhand des Jonckheere-Terpstra Tests (Trendhypothese: Profil 1 > 2 > 3 [> 4]). Angegeben ist die standardisierte Prüfgrösse.

* Das Ergebnis ist auf dem 5%-Niveau signifikant.

** Das Ergebnis ist auf dem 1%-Niveau signifikant.

*** Das Ergebnis ist auf dem 0.1%-Niveau signifikant.

Auch der Zusammenhang zwischen den *Prozessprofilen* und den Leistungskennwerten kann bestätigt werden und geht mehrheitlich in die erwartete Richtung (Hypothese 4.2). Sowohl die drei Prozessprofile für das Informationseröffnende Handeln wie auch die vier Prozessprofile für das Adäquate Handeln unterscheiden sich *signifikant in neun der zehn geprüften Kennwerte* (vgl. Tabelle 6). Wiederum kann für den Kennwert „Bearbeitungszeit für eingegangene Informationen effizient nutzen (p_ow_m)“ kein signifikanter Zusammenhang nachgewiesen werden.

Beim *Informationseröffnenden Handeln* kann die postulierte *Trendhypothese* (der Median in Profil 1 ist grösser als in Profil 2 und dieser wiederum ist grösser als in Profil 3) *für alle Leistungskennwerte* bestätigt werden. Auch wenn die Gruppenunterschiede für den Kennwert „Bearbeitungszeit für eingegangene Informationen effizient nutzen

(p_{ow_m})“ nicht signifikant sind, so entspricht die Reihenfolge der Medianwerte doch der erwarteten Richtung. Abbildung 13 zeigt den Ausprägungsgrad der Mediane in den Leistungskennwerten in Abhängigkeit von den Prozessprofilen. Die exakten Medianwerte sind in Anhang D aufgeführt (Tabelle 15).

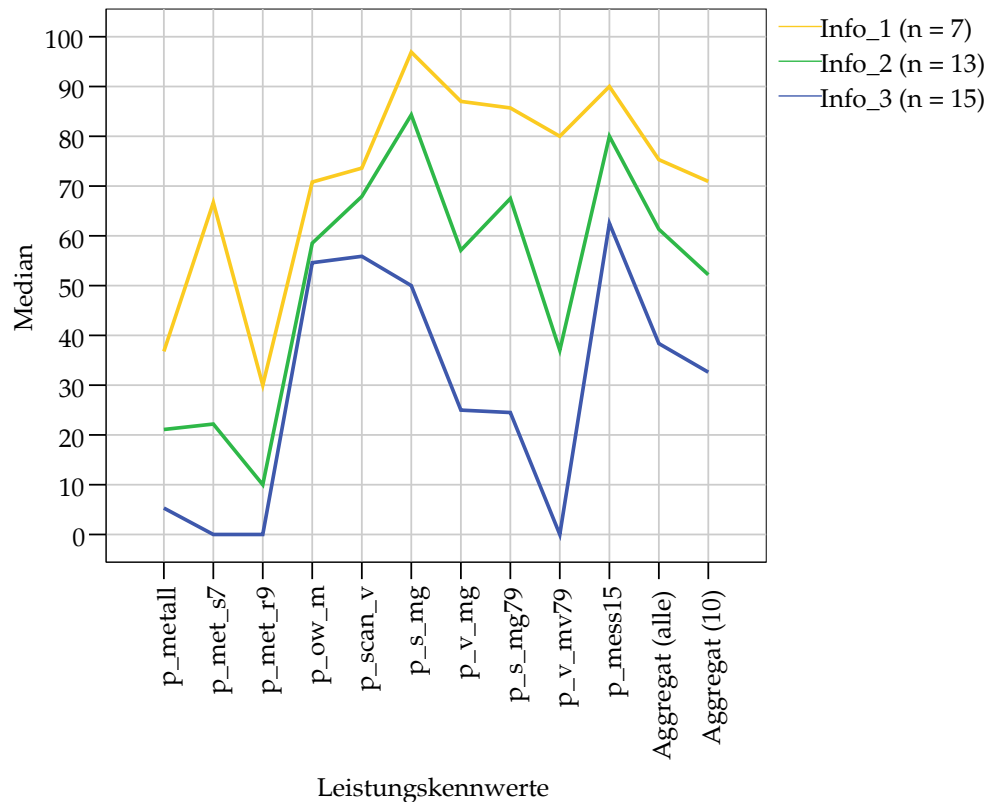


Abbildung 13: Liniendiagramm der Mediane in den szenariospezifischen Leistungskennwerten für die Prozessprofile des Informationseröffnenden Handelns (n = 35, für Aggregat über alle Kennwerte ist n = 28).

Beim *Adäquaten Handeln* kann die erwartete Reihenfolge der Medianwerte (Profil 1 > 2 > 3 > 4) nur für vier Leistungskennwerte bestätigt werden: für den Kennwert „Effiziente Verschiebung der Suchpunkte (p_{scan_v})“, den Kennwert „Suchaktivitäten in sinnvollen Gebieten (p_{s_mg})“, den Kennwert „Prüfaktivitäten in sinnvoller Tiefe (p_{v_mg})“ sowie für den Kennwert „Suchaktivitäten in sinnvoller Tiefe und sinnvollen Gebieten (p_{s_mg79})“. Für die Kennwerte p_{s_mg} sowie p_{s_mg79} sind dabei die Unterschiede zwischen den Profilen 1 und 2 nur gering (vgl. Abbildung 14 und Tabelle 15 in Anhang D).

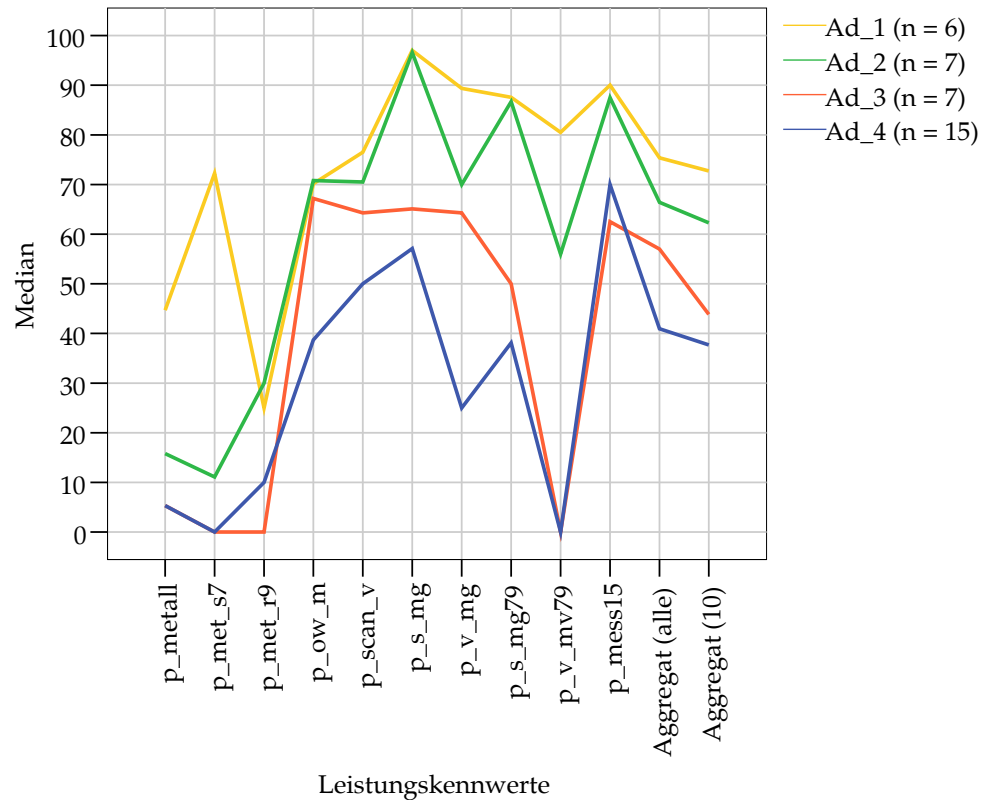


Abbildung 14: Liniendiagramm der Mediane in den szenariospezifischen Leistungskennwerten für die Prozessprofile des Adäquaten Handelns (n = 35, für Aggregat über alle Kennwerte ist n = 28).

Bei drei weiteren Kennwerten zeigt sich zwar ein monoton fallender Trend, jedoch *ohne Unterschied in den Profilen 3 und 4*: beim Kennwert „Anteil gescannter Fläche in sinnvollen Gebieten (*p_metall*)“, beim Kennwert „Anteil gescannter Fläche in S/7 (*p_met_s7*)“ sowie beim Kennwert „Prüfaktivitäten in sinnvoller Tiefe und sinnvollen Gebieten (*p_v_mv79*)“.

Kein monoton fallender Trend lässt sich für die drei Kennwerte *p_met_r9*, *p_ow_m* und *p_mess15* nachweisen. Die Reihenfolge der Profile ist dabei wie folgt: $2 > 1 > 4 > 3$ für den Kennwert „Anteil gescannter Fläche in R/9 (*p_met_r9*)“, $2 > 1 > 3 > 4$ für den Kennwert „Bearbeitungszeit für eingegangene Informationen effizient nutzen (*p_ow_m*)“ und $1 > 2 > 4 > 3$ für den Kennwert „Ankommende Informationen vollständig abrufen (*p_mess15*)“. Die Unterschiede zwischen den Medianwerten sind bei diesen drei Kennwerten zudem geringer als bei den anderen Kennwerten (und für *p_ow_m* nicht signifikant).

Die Prozessprofile der beiden Konstrukte unterscheiden sich auch signifikant im Hinblick auf die beiden *aggregierten Masse* (vgl. Tabelle 6). Die Reihenfolge der Medianwerte entspricht für beide Masse und beide Konstrukte eindeutig der erwarteten Richtung (vgl. Abbildungen 13 und 14 sowie Tabelle 15 in Anhang D).

4.4.2 Vergleich mit dem Expertenurteil

Die Einschätzungen der Probandenleistung durch das neue Auswertungsmodell und den Experten korrelieren hoch und signifikant (Info: $r_s = .70$, $p < .01$, Ad: $r_s = .71$, $p < .01$).

In Bezug auf die Beurteilung der Urteilsübereinstimmung zeigt sich, dass die Übereinstimmung für das Informationseröffnende Handeln etwas besser ist als für das Adäquate Handeln. Tabelle 7 und Tabelle 8 zeigen die entsprechenden Kontingenztafeln. Die Kategorien 1 bis 3 (Info) und 1 bis 4 (Ad) entsprechen der Reihenfolge der Prozessprofile. Bei beiden Urteilen werden in etwa gleich grosse Gruppen gebildet. Insgesamt stimmen beim Informationseröffnenden Handeln 69% der Urteile überein (24 Probanden), beim Adäquaten Handeln 54% der Urteile (19 Probanden). Die Diskordanzen betreffen vor allem *benachbarte* Kategorien. Abweichungen von der schlechtesten zur besten Urteilstkategorie kommen bei beiden Konstrukten nicht vor, und bei beiden Konstrukten sind die Abweichungen im *unteren* Leistungsbereich (Profil 2-3 bzw. 3-4) grösser als im oberen Bereich (Profil 1-2).

Tabelle 7: Übereinstimmung zwischen neuem Auswertungsmodell und Expertenurteil für das Informationseröffnende Handeln in der Teilstichprobe 1.

		Neues Auswertungsmodell			Total
		1	2	3	
Expertenrating	1	5 (14%)	1 (3%)	0	6
	2	2 (6%)	7 (20%)	3 (9%)	12
	3	0	5 (14%)	12 (34%)	17
Total		7	13	15	35

Tabelle 8: Übereinstimmung zwischen neuem Auswertungsmodell und Expertenurteil für das Adäquate Handeln in der Teilstichprobe 1.

		Neues Auswertungsmodell				Total
		1	2	3	4	
Expertenrating	1	5 (14%)	1 (3%)	0	0	6
	2	1 (3%)	2 (6%)	1 (3%)	0	4
	3	0	3 (9%)	2 (6%)	5 (14%)	10
	4	0	1 (3%)	4 (11%)	10 (29%)	15
Total		6	7	7	15	35

Beim Informationseröffnenden Handeln werden bei drei Probanden (9%) die Kategorien 1 und 2 vertauscht, bei acht Probanden (23%) die Kategorien 2 und 3. Beim Adäquaten Handeln betreffen die Diskordanzen bei zwei Probanden (6%) die Kategorien 1-2, bei vier Probanden (11%) die Kategorien 2-3 und bei neun Probanden (26%) die Kategorien 3-4. Nur bei einem Probanden (3%) betrifft die Diskordanz nicht benachbarte Kategorien (Abweichung 2-4).

Damit ergibt sich für das Informationseröffnende Handeln ein *gewichteter Kappa-Koeffizient* (κ_W) von .70 und für das Adäquate Handeln von .60 (für die Festlegung der Diskrepanzgewichte vgl. Kapitel 3.5). Für beide Werte kann eine Irrtumswahrscheinlichkeit von 0.1% angenommen werden. Aufgrund der kleinen Stichprobengrösse und insbesondere der ungleichen Verteilung der Probanden auf die Urteilkategorien, sind die e_{ij} -Werte (theoretisch erwartete Häufigkeit der Urteilsübereinstimmungen in Zelle ij , wenn die Urteile stochastisch unabhängig wären) teilweise kleiner als zwei, womit die Voraussetzung für eine Signifikanzbeurteilung des beobachteten κ_W verletzt ist. Die Signifikanzangabe darf deshalb nur als Schätzwert betrachtet werden.

Werden die beiden Urteile nicht in Bezug auf die Prozessprofile der beiden Konstrukte, sondern in Bezug auf das aus diesen resultierende *Beurteilungsschema* verglichen, so stimmen insgesamt nur 40% der Urteile (14 Probanden) überein. Diese

geringe Übereinstimmung ist darauf zurückzuführen, dass die Diskordanzen bei den Prozessprofilen des Informationseröffnenden und Adäquaten Handelns mehrheitlich nicht die gleichen Probanden betreffen: Sechs Probanden (17%) werden bei beiden Konstrukten unterschiedlichen Prozessprofilen zugeordnet, während bei 15 Probanden (43%) die Diskordanzen entweder das eine oder das andere Konstrukt betreffen. Wie schon aus den Resultaten zur Übereinstimmung der Prozessprofile hervorgeht, betreffen jedoch alle Diskordanzen – mit einer Ausnahme – (horizontal, vertikal oder diagonal) *benachbarte* Problemlösetypen. Wie beim neuen Auswertungsmodell kommen auch beim Expertenurteil die Kombinationen aus dem jeweils besten und dem (zweit-) schlechtesten Profil (I1-A3, I1-A4, A1-I3) nicht vor.

Für den im Rahmen der Personalauswahl interessierenden *oberen Leistungsbereich* ist die Übereinstimmung sehr gut: Alle neun Probanden (26%), die vom Experten den guten bis sehr guten Problemlösern zugeordnet werden (obere vier Felder 1-1, 1-2, 2-1, 2-2), werden auch beim neuen Auswertungsmodell diesen Feldern zugeteilt. Nach dem neuen Auswertungsmodell gehören noch drei weitere Probanden zu den guten Problemlösern (zum Typ I2-A2). Diese werden vom Experten den Typen I2-A4 (Nr. 13), I3-A3 (Nr. 26) sowie I2-A3 (Nr. 35) zugeordnet. Die Diskordanzen im oberen Leistungsbereich betreffen damit nur das Adäquate Handeln. Detaillierte Angaben zu den Konkordanzen für einzelne Problemlösetypen sind Anhang D zu entnehmen (Tabelle 16).

4.4.3 Überprüfung an einem unabhängigen Datensatz

Die Verhaltensprotokolle der 21 Probanden der Teilstichprobe 2 liessen sich problemlos entsprechend den an der Teilstichprobe 1 entwickelten Interpretationsregeln der Mikroverhaltensbeschreibung codieren. Auch konnten alle Probanden durch Anwendung der gewählten Klassifikationsregeln eindeutig den Prozessprofilen für das Informationseröffnende und Adäquate Handeln zugeordnet werden (Hypothese 6.1).

Auch in der Teilstichprobe 2 zeigt sich ein signifikanter Zusammenhang zwischen den Einschätzungen der Probandenleistung durch das neue Auswertungsmodell

und den Einschätzungen durch den Experten (Hypothese 6.2). Der Zusammenhang ist allerdings für beide Konstrukte etwas geringer als in der Teilstichprobe 1 (Info: $r_s = .55$, $p < .05$, Ad: $r_s = .52$, $p < .05$). Gemäss der Effektstärkenklassifikation von Cohen (1992, S. 157) liegen aber immer noch grosse Effekte vor.

Wie Tabelle 9 und Tabelle 10 zeigen, bilden Experte und neues Auswertungsmodell beim Informationseröffnenden Handeln – analog zur Teilstichprobe 1 – in etwa gleich grosse Gruppen. Für das Adäquate Handeln weichen die Gruppengrössen jedoch für die Kategorien 2 und 4 deutlich voneinander ab (Experte/Auswertung: 8/3 in Kategorie 2 und 6/11 in Kategorie 4). In Bezug auf den Anteil der konkor-danten Urteile zeigen sich nur geringe Unterschiede zur Teilstichprobe 1. Beim Informationseröffnenden Handeln stimmen 62% *der Urteile* überein (13 Probanden), beim Adäquaten Handeln 52% (11 Probanden). Wiederum betreffen die Diskordan-zen vor allem *benachbarte* Kategorien, und Diskordanzen zwischen den besten und schlechtesten Kategorien kommen nicht vor.

Beim Informationseröffnenden Handeln sind die Diskordanzen im Unterschied zur Teilstichprobe 1 im oberen und unteren Leistungsbereich gleich verteilt. Die Abwei-chungen betreffen bei je vier Probanden (je 19%) die Kategorien 1-2 sowie 2-3. Beim Adäquaten Handeln betreffen die Diskordanzen bei einem Probanden (5%) die Kate-gorien 1-2, bei zwei Probanden (10%) die Kategorien 2-3 und bei sieben Probanden (33%) die Kategorien 3-4. Bei *vier* Probanden (19%) betrifft die Diskordanz *nicht benachbarte* Kategorien (Abweichung 2-4).

Tabelle 9: Übereinstimmung zwischen neuem Auswertungsmodell und Expertenurteil für das Informationseröffnende Handeln in der Teilstichprobe 2.

		Neues Auswertungsmodell			Total
		1	2	3	
Expertenrating	1	0	3 (14%)	0	3
	2	1 (5%)	8 (38%)	2 (10%)	11
	3	0	2 (10%)	5 (24%)	7
Total		1	13	7	21

Tabelle 10: Übereinstimmung zwischen neuem Auswertungsmodell und Expertenurteil für das Adäquate Handeln in der Teilstichprobe 2.

		Neues Auswertungsmodell				Total
		1	2	3	4	
Expertenrating	1	1 (5%)	1 (5%)	0	0	2
	2	0	2 (10%)	2 (10%)	4 (19%)	8
	3	0	0	3 (14%)	2 (10%)	5
	4	0	0	1 (5%)	5 (24%)	6
Total		1	3	6	11	21

Im Vergleich zur Teilstichprobe 1 ist demnach die relative Häufigkeit der stärker gewichteten Diskordanzen grösser (Info: 19% versus 9% in Kategorie 1-2; Ad: 19% versus 3% in Kategorie 2-4), so dass die *gewichteten Kappa-Koeffizienten* deutlich schlechter ausfallen (Info: $\kappa_W = .34$, $p < .01$, Ad: $\kappa_W = .40$, $p < .05$; die Irrtumswahrscheinlichkeiten sind wiederum nur als Schätzwerte zu verstehen).

In Bezug auf die Einordnung im *Beurteilungsschema* stimmen 38% der Urteile (acht Probanden) überein, was wiederum darauf zurückzuführen ist, dass die Diskordanzen in den Prozessprofilen der beiden Konstrukte zu einem grossen Teil nicht die gleichen Probanden betreffen: Fünf Probanden (23%) werden in beiden Konstrukten unterschiedlich beurteilt, acht (38%) nur im einen oder anderen Konstrukt. Die Kombinationen aus dem jeweils besten und dem (zweit-) schlechtesten Profil (I1-A3, I1-A4, A1-I3) kommen auch in der Teilstichprobe 2 nicht vor.

Für den im Rahmen der Personalauswahl interessierenden *oberen Leistungsbereich* zeigt sich, dass das neue Auswertungsmodell konservativer ist als der Experte: Nur vier Probanden (19%) werden den vier besten Problemlösetypen (Felder 1-1, 1-2, 2-1, 2-2) zugeordnet, während dies beim Experten neun (43%) sind. Immerhin werden die gleichen vier Probanden von beiden als gute bis sehr gute Problemlöser beurteilt. Detaillierte Angaben zu den Konkordanzen für einzelne Problemlösetypen sind wiederum Anhang D zu entnehmen (Tabelle 17).

5 Diskussion und Ausblick

In dieser Arbeit wurde ein neues Auswertungsmodell für computergestützte Problemlöseszenarien zum Zwecke der Personalauswahl vorgestellt. Ausgangspunkt der Entwicklung des neuen Modells war die Feststellung, dass Problemlöseszenarien innerhalb der Eignungsdiagnostik immer noch relativ wenig verbreitet sind, obwohl viele Gründe für deren Anwendung in diesem Kontext sprechen. Neben der heterogenen Befundlage im Hinblick auf die eignungsdiagnostische Brauchbarkeit der Szenarien wurde insbesondere die mangelnde Standardisierung der Auswertung, vor allem der Dateninterpretation und -integration, dafür verantwortlich gemacht. Dies sind auch die Hauptgründe dafür, dass einige Autoren die weitere Verwendung der Szenarien im Rahmen der Personalselektion kritisch beurteilen (J. Funke, 1998; Kersting, 1999a, 1999b, 2001; Kersting & Kluge, 2001; Kleinmann & Strauss, 1998; Süß, 1996, 1999).

Im Unterschied dazu wurde in dieser Arbeit argumentiert, dass bestehende Auswertungsansätze das Potential von Problemlöseszenarien noch nicht vollständig ausgeschöpft haben. So ergab die Zusammenstellung der vorliegenden Ansätze, dass die Möglichkeit der Szenarien zur *Prozessdiagnostik* – was immer wieder als der entscheidende Vorteil gegenüber herkömmlichen Leistungstests aufgeführt wird – nur selten, und wenn überhaupt dann meistens einseitig, genutzt wird. Als prozessorientiert bezeichnete Ansätze berechnen vorwiegend *globale* Masse, indem *ausgewählte* Verhaltens- oder Wissensaspekte ausgezählt und bewertet werden (für Ausnahmen vgl. Strohschneider, 1994; Wirth, 2004; Wirth & Funke, 2005). Prozessorientierung bei Problemlöseszenarien kann jedoch mehr bedeuten: die Beurteilung *jedes einzelnen Handlungsschrittes* (d.h. aller automatisch registrierbaren Aktionen, in der Regel Mausklicks, bezeichnet als Mikroverhalten) *im aktuellen Handlungskontext*, und die Beurteilung des *Problemlöseprozesses als Ganzes*.

Ein weiterer Kritikpunkt betraf die Feststellung, dass in jedem Szenario andere Indikatoren der Problemlöseleistung herangezogen werden, die zudem (vorwiegend) *szenariospezifisch* sind, das heisst nur für dieses Szenario beziehungsweise für Szenarien mit dem gleichen Inhalt berechnet werden können. Die sich daraus ergebende

Vielfalt an Problemlösegütemassen erschwert eine vergleichende Interpretation von Befunden.

Das neue Auswertungsmodell berücksichtigt diese Schwachpunkte. Ergebnis der Auswertung ist ein automatisierbares Beurteilungsschema, anhand dessen der Diagnostiker schnell und ohne grosse Vorkenntnisse das gezeigte Problemlöseverhalten bewerten kann. Das Beurteilungsschema basiert auf

- einer prozessorientierten Mikroverhaltensbeschreibung, bei der *jedes Mikroverhalten* (jeder Handlungsschritt) *im aktuellen Handlungskontext beurteilt* wird (im Hinblick auf zwei Konstrukte, die für das Lösen komplexer Probleme grundlegend sind: *Informationseröffnendes und Adäquates Handeln*, abgekürzt mit Info und Ad),
- der graphischen Darstellung der zeitlichen Entwicklung des Informationseröffnenden und Adäquaten Handelns in Form von *Zeitreihen*,
- der Beschreibung der Zeitreihen durch *Prozessmasse* und so genannte *Prozessprofile*, deren Definition auf der *Klassifizierung der Zeitreihen* basiert und die eine *Bewertung der Problembearbeitung* beinhalten, sowie
- der Beschreibung von *Problemlösetypen* (typischen Bearbeitungsmustern), die sich aus der Kombination der Prozessprofile der beiden Konstrukte ergeben.

Damit beinhaltet das Beurteilungsschema nicht nur eine Standardisierung der Datenauswertung, sondern auch eine Standardisierung der Dateninterpretation und Datenintegration. Resultat ist ein vom Inhalt des jeweiligen Szenarios unabhängiges (szenario-unspezifisches), mikroverhaltensbasiertes, prozessorientiertes Gesamturteil über die gezeigte Problemlöseleistung.

Ziel dieser Arbeit war, dieses Auswertungsmodell auf ein Problemlöseszenario anzuwenden und erste Aussagen über die Gültigkeit der Auswertung zu machen. Die Wahl eines geeigneten Problemlöseszenario fiel auf das Szenario ATLANTIS (Hirsig & de With, 1996, 1998, 2001), weil dieses die gestellten Anforderungen an *Szenarienkonstruktion* (vgl. Hasselmann, 1995; Kröner et al., 2005), *Datenspeicherung* (zeitsynchron, Mikroverhalten und Handlungskontext) und *Anwendung* (Nachweis eines „erfolgreichen“ Einsatzes in Personalselektion und -entwicklung) erfüllt.

An einer ersten Teilstichprobe wurden die Interpretationsregeln der Mikroverhaltensbeschreibung entwickelt und das Beurteilungsschema definiert. Anhand der

Darstellung und Beschreibung der 35 Zeitreihenpaare liess sich zeigen, dass das neue Auswertungsmodell die individuellen Unterschiede im Problemlöseverhalten zu beschreiben vermag (Fragestellung 1). Die Verläufe für das Informationseröffnende und das Adäquate Handeln variieren bei den 35 Probanden beträchtlich, wobei die Varianz in den Ereignispunkten mit dem Spielverlauf zunimmt. Die Verläufe unterscheiden sich neben der *Verlaufsform* (steil und konstant ansteigend, flach ansteigend, degressiv ansteigend (d.h. abflachend), um Null schwankend, abfallend, Mischformen) auch in Merkmalen wie der Anzahl und Länge von *Plateaus* oder der Häufigkeit und Höhe von *Spitzen* beziehungsweise Wechsel im Verlauf (nur bei Ad, Hinweis auf Inkonsistenz des Verhaltens). Wesentliches Merkmal im Gesamtbild ist das *Auseinanderklaffen der beiden Kurven* und die separate Betrachtung der *Verläufe während der Trainings- und der Testphase* (durch Rekalibrierung am Trainingsendwert können die beiden Verläufe für die Testphase eine ganz andere inhaltliche Bedeutung erhalten).

Letzteres ist deshalb wichtig, weil aufgrund mangelnder Standardisierung des Spielverlaufs während dem Training die Probanden nicht die gleichen Vorbedingungen haben (zur Erinnerung: Die Trainingszeiten variieren zwischen 17 Minuten und knapp einer Stunde!). Das heisst, das Wissen, das sich die Probanden während dieser Phase aneignen und das sie für ein erfolgreiches Lösen der Testaufgabe benötigen, ist nicht direkt vergleichbar. Da im neuen Auswertungsmodell aber gerade interessiert, inwiefern sich die Probanden mit jedem Handlungsschritt neue Informationen eröffnen und damit potientiell Wissen aufbauen, sollte bei einer Weiterentwicklung des Szenarios eine Standardisierung der Übungsphase unbedingt angestrebt werden (analog zu Ansätzen der Wissenserwerbsdiagnostik, z.B. Kröner et al., 2005; Wirth & Funke, 2005).

Durch Interpretation und Bewertung der Verlaufsformen und Merkmale für die Problemlösung liessen sich für das Informationseröffnende Handeln drei *Prozessprofile* definieren, für das Adäquate Handeln (aufgrund der grösseren Varianz der Verläufe) vier (Fragestellung 2). Während die Prozessprofile 1 bei beiden Konstrukten homogene Gruppen darstellen, bestehen die übrigen Profile aus jeweils unterschiedlichen Verlaufsformen, die in Bezug auf die Güte der Problemlösung

jedoch die gleiche Bedeutung haben.

Die inhaltlich orientierte, etwas willkürliche Klassifizierung der Zeitreihen liess sich durch *formale Klassifizierungsmethoden* recht gut bestätigen (Fragestellung 3), wobei vor allem die Klassifizierung der jeweils besten Probanden (Prozessprofile 1) sehr gut gelang. Zwischen den *unterschiedlichen Ähnlichkeitsformalisierungen* (A: über das Prozessmass, B: über den Abstand zwischen den Kurven in Form Euklidischer Distanzen; drei Berechnungsvarianten) und *unterschiedlichen Klassifizierungsmethoden* (verschiedene Algorithmen der Clusteranalyse, Multidimensionale Skalierung) ergaben sich nur geringfügige Unterschiede, die zudem meist andere Probanden betrafen. Die wenigen Abweichungen von der inhaltlich orientierten Klassifizierung liessen sich mehrheitlich durch Methodenartefakte erklären. Die Ergebnisse zur Klassifizierung der Zeitreihen wurden im Kapitel 4.3.2 ausführlich interpretiert und diskutiert, ebenso die aus den Prozessprofilen resultierenden Problemlösetypen (vgl. Kapitel 4.3.3).

Die *Überprüfung des neuen Auswertungsmodells* ergab, dass dieses die Problemlöseleistung ähnlich beurteilt wie der *szenariospezifische Auswertungsansatz*, der während mehrerer Jahre im Bereich der Personalauswahl und -entwicklung angewendet wurde (Fragestellung 4). Auch ist die Beurteilung vergleichbar mit den Einschätzungen eines *Experten*, der die verschiedenen, zum Teil widersprüchlichen Ergebnisse der szenariospezifischen Auswertung zu einem dem Auswertungsmodell analogen, szenario-unspezifischen Gesamturteil integriert (Fragestellung 5). Ein Zusammenhang zwischen neuem Auswertungsmodell und Expertenurteil liess sich auch an einem von der Entwicklung des Auswertungsmodells *unabhängigen Datensatz* (Teilstichprobe 2) nachweisen (Fragestellung 6). Die Übereinstimmung zwischen den beiden Urteilen war allerdings etwas geringer als in der Teilstichprobe 1, dennoch kann damit ausgeschlossen werden, dass die definierten Interpretations- und Klassifikationsregeln nur für die „Entwicklungs-Stichprobe“ Gültigkeit haben. Die Ergebnisse zu diesen drei Fragestellungen werden im Folgenden zusammengefasst und diskutiert.

Vergleich mit der szenariospezifischen Auswertung

Für diesen Vergleich wurden die 13 Leistungskennwerte (Verhaltensmasse) gewählt, da diese auch in der diagnostischen Praxis Hauptbestandteil der Urteilsbildung sind und zudem eine Art Normwerte dafür vorliegen (vgl. dazu Kapitel 3.1.3.3). Drei dieser Leistungskennwerte mussten aufgrund mangelnder Differenzierung (Median über 90, geringe Interquartilweite) ausgeschlossen werden (p_{ow_i} , $p_{m_p_nv}$, p_{iseit} , vgl. Beschreibung in Tabelle 2, Seite 63).

Sowohl die beiden Prozessmasse wie auch die Prozessprofile stehen mit *neun der zehn geprüften Leistungskennwerte* in einem signifikanten Zusammenhang. Einzig für den Kennwert „Bearbeitungszeit für eingegangene Informationen effizient nutzen (p_{ow_m})“ liess sich kein Zusammenhang zwischen den beiden Auswertungen nachweisen. Der Kennwert beschreibt, wie häufig eingegangene Meldungen wiederholt gelesen werden (Lesezeit für nicht wiederholte Seiten in Relation zur Gesamtlesezeit), wobei das häufige Wiederholen negativ bewertet wird. Im Unterschied dazu werden beim neuen Auswertungsmodell Wiederholungen nicht durchgehend negativ bewertet. Zwar wäre es für eine effiziente Problemlösung wichtig, bereits beim ersten Lesen die gesamte Information aufzunehmen, dennoch kann eine *erste* Wiederholung, wenn auch wenig sinnvoll ($Ad = 0$), im Sinne von „beim ersten Mal etwas übersehen haben“ noch ein Mehr an relevanter Information bedeuten ($Info = +0.5$). Erst jede weitere Wiederholung wird, sofern die minimale Lesedauer erfüllt ist, negativ bewertet ($Info = 0$, $Ad = -1$). Diese unterschiedliche Bewertung von Wiederholungen beim Lesen könnte ein Grund für den nicht signifikanten Zusammenhang sein.

Die Korrelationen zwischen den *Prozessmassen* (Info- und Ad-Endwerte) und den Leistungskennwerten liegen zwischen .31 und .75. Bei den Leistungskennwerten handelt es sich um spezifische Verhaltensskalen, die nicht auf dem gleichen Generalitätsniveau vorliegen wie die Prozessmasse und Prozessprofile. Deshalb wurden – in Anlehnung an die multivariate Reliabilitäts- und Validitätstheorie von Wittmann (1987) und Wittmann und Matt (1986) – aus den Leistungskennwerten zwei aggregierte Masse gebildet (alle 13 Kennwerte, nur die zehn gut differenzierenden Kennwerte). Auf dieser Ebene ergaben sich sogar Korrelationen über .84.

Für die *Prozessprofile* konnte die postulierte *Trendhypothese* über die Reihenfolge der Medianwerte für das Informationseröffnende Handeln für alle zehn Leistungskennwerte und beide aggregierte Masse bestätigt werden. Auch wenn die Gruppenunterschiede für den Kennwert *p_ow_m* nicht signifikant waren, so haben Profil-1-Probanden dennoch tendenziell höhere Werte als Profil-2-Probanden und diese wiederum höhere als Profil-3-Probanden. Beim Adäquaten Handeln zeigte sich der monoton fallende Trend in sieben Leistungskennwerten (wovon bei drei Kennwerten Profil 3 und 4 die gleichen Medianwerte haben) sowie bei den aggregierten Massen. Bei den drei Kennwerten *p_met_r9*, *p_ow_m* und *p_mess15* lagen die Medianwerte der Profile 1 und 2 und/oder der Profile 3 und 4 in umgekehrter Reihenfolge vor. Immerhin haben die besseren Probanden (Profil 1 und 2) auch in diesen Kennwerten die höheren Medianwerte als die schlechteren Probanden (Profil 3 und 4). Ausserdem handelt es sich bei *p_ow_m* und *p_mess15* um Kennwerte, die für die Beurteilung der Problemlöseleistung nach Angabe der Testautoren⁹ eine geringere Bedeutung haben. Besonders relevant sind demnach die vier Kennwerte der Problemerkennung und -lösung sowie die drei Kennwerte des Agierens, wobei bei letzteren *p_met_r9* ebenfalls mehr ergänzend zu verstehen ist (als Hinweis auf die Flexibilität, d.h. die Frage, ob auch ein Wechsel der Suchgebiete stattgefunden hat).

Vergleich mit dem Expertenurteil

In der Teilstichprobe 1 ergab sich eine Übereinstimmung von 69% (Info) beziehungsweise 54% (Ad) der Urteile. Mit Korrelationen von $r_s \geq .70$ liegen starke Effekte vor, und die gewichteten Kappa-Koeffizienten (κ_w) von .70 (Info) und .60 (Ad) sind ebenfalls zufrieden stellend. Bei beiden Konstrukten betreffen die Diskordanzen (mit einer Ausnahme) *benachbarte* Urteilkategorien (Prozessprofile), und die Abweichungen im *unteren* Leistungsbereich sind deutlich grösser als im *oberen* Leistungsbereich (Info: 23% in Kategorien 2-3 versus 9% in 1-2, Ad: 37% in 2-3 und 3-4 versus 6% in 1-2). Auch im Beurteilungsschema, das sich aus den Kombinationen der Prozessprofile der beiden Konstrukte ergibt, zeigt sich für den oberen, für die Personalauswahl relevanten Leistungsbereich eine hohe Übereinstimmung: Alle

⁹ Persönliche Mitteilung der Testautorin Dr. Astrid E. de With; vgl. auch Kapitel 3.1.3.3.

neun vom Experten als gute bis sehr gute Problemlöser beurteilten Probanden werden auch beim neuen Auswertungsmodell diesem Leistungsbereich (Info-Ad-Felder 1-1, 1-2, 2-1, 2-2) zugeordnet.

In der *Teilstichprobe 2* zeigte sich eine etwas schlechtere Übereinstimmung zwischen den beiden Einschätzungen. Der relative Anteil der konkordanten Urteile (Info: 62%, Ad: 52%) ist zwar annähernd so hoch wie in der Teilstichprobe 1, der relative Anteil der stärker gewichteten Diskordanzen (Info-Kategorien 1-2 und Ad-Kategorien 2-4) ist jedoch deutlich höher, weshalb die gewichteten Kappa-Koeffizienten geringer ausfallen ($\kappa_W = .34$ für Info und $.40$ für Ad). Mit Korrelationen über $.51$ liegen aber immer noch starke Effekte vor. In Bezug auf die Problemlösetypen werden die gleichen vier Probanden vom Experten und dem neuen Auswertungsmodell als die besten beurteilt.

Die *Unterschiede zwischen Expertenurteil und neuem Auswertungsmodell* lassen sich teilweise anhand von Überlegungen zur inhaltlich orientierten Klassifizierung erklären. Bei einigen Abweichungen handelt es sich um *Grenzfälle*, das heisst Probanden, die beim neuen Auswertungsmodell die Einordnung ins höhere Profil nur knapp um wenige Punkte verpassen oder umgekehrt nur knapp in ein höheres Profil „reinrutschen“. Einige wenige Beispiele sollen dies veranschaulichen:

- Alle drei Probanden der Teilstichprobe 1, die nach dem neuen Auswertungsmodell dem *Info-Profil 3* und nach dem Experten der *Info-Kategorie 2* zugeordnet werden (Nr. 33, 21 und 25), erfüllen die Klassifikationskriterien für Profil 2 nur ganz knapp nicht. Während bei Proband Nr. 33 beide Kriterien um 2.5 beziehungsweise 3 Punkte nicht erfüllt sind, ist bei den beiden anderen nur jeweils ein Kriterium um 3.5 Punkte (Nr. 21) beziehungsweise sogar um nur einen Punkt (Nr. 25) nicht erfüllt.
- Umgekehrt wird Proband Nr. 31 beim neuen Auswertungsmodell ebenfalls sehr knapp (wegen einem halben Punkt) dem *Info-Profil 2* zugeteilt (statt wie beim Experten der *Info-Kategorie 3*).
- In der *Teilstichprobe 2* handelt es sich bei einigen Grenzfällen sogar genau um die stärker gewichteten Diskordanzen, die für die niedrigere Konkordanz (gewichteter Kappa-Koeffizient) verantwortlich sind: Drei der vier Probanden mit Diskordanzen in den *Info-Kategorien 1-2* (Nr. 6, 9, 20) verpassen die Einordnung ins Info-Profil 1 ebenfalls um wenige Punkte in einem Kriterium, während das zweite Kriterium bei allen erfüllt ist.

Beim Adäquaten Handeln lassen sich ebenfalls einige Grenzfälle feststellen. Diese sind vor allem auf die *unterschiedliche Gewichtung von Trainings- und Testphase*

zurückzuführen. Beim neuen Auswertungsmodell wurde bei der Klassifizierung der Zeitreihen der Verlauf der Testphase nicht nur mehr gewichtet als der Verlauf während dem Training, sondern auch separat betrachtet (durch Rekalibrierung am Trainingsendwert). Beim Adäquaten Handeln hat diese Gewichtung aufgrund der grösseren Vielfalt der Verlaufsformen einen stärkeren Einfluss auf die Klassifizierung als beim Informationseröffnenden Handeln. Auch hierfür ein paar Beispiele:

- Die Probanden Nr. 13 und Nr. 33 werden beim neuen Auswertungsmodell ins *Ad-Profil 2* eingeordnet, obwohl der Verlauf während dem Training flach oder sogar abfallend ist (Profil 2.2 mit Startschwierigkeiten; vgl. auch Bsp. B und C in Abbildung 5, Seite 98). In Bezug auf die gesamte Problembearbeitung sind sie nicht mit Profil-2-Probanden vergleichbar, in Bezug auf die Leistung während der Testphase jedoch schon. Da der Experte die gesamte Bearbeitungszeit berücksichtigt, teilt er die beiden den *Ad-Kategorien 3* (Nr. 33) beziehungsweise 4 (Nr. 13) zu.
- Ein anderes Beispiel ist Proband Nr. 1 mit seinem wechselnden Verlauf (vgl. Bsp. E in Abbildung 5, Seite 98). Gesamthaft ist der Verlauf, wenn auch stark schwankend, mehrheitlich positiv, was für eine Einordnung ins *Ad-Profil 3* sprechen würde. Wird nur die Testphase betrachtet, verläuft die Kurve jedoch mehrheitlich im negativen Bereich und der Proband zeigt keine Leistungssteigerung, weshalb er beim neuen Auswertungsmodell dem *Ad-Profil 4* zugeordnet wird (beim Experten der *Ad-Kategorie 3*).
- Ebenfalls zu dieser Art Grenzfälle gehören die beiden in Kapitel 4.3.3 näher beschriebenen Probanden Nr. 2 (Teilstichprobe 1) und Nr. 5 (Teilstichprobe 2), deren *Ad-Verlauf* erst nach Erhalt der Notfallmeldungen M3 beziehungsweise M2 abfällt, während der Verlauf davor auch eine Einordnung ins *Ad-Profil 3* erlauben würde (analog zum Urteil des Experten).

Das Argument der Grenzfälle ist allerdings nur bedingt geeignet, die Unterschiede zwischen Expertenurteil und neuem Auswertungsmodell zu erklären. Einerseits trifft es nur auf einen Teil der Diskordanzen zu, und andererseits lassen sich auch Grenzfälle bei der inhaltlich orientierten Klassifizierung finden, die konkordante Urteile betreffen. Bei jeder Klassifizierung gibt es Probanden, die die Klassifikationsregeln nur knapp erfüllen beziehungsweise nicht erfüllen, das heisst Probanden, die anhand der gebildeten Klassen schlecht beschrieben werden können. Jede Klassifizierung ist in grossem Masse *stichprobenabhängig*. Dass die Übereinstimmung zwischen Expertenurteil und neuem Auswertungsmodell in der von der Entwicklung der Klassifizierung unabhängigen Teilstichprobe 2 schlechter ausfiel als in der Teilstichprobe 1, könnte ein Hinweis darauf sein, dass die vorgenommene Klassifizierung noch optimiert werden kann. Beide Stichproben waren relativ klein. Für eine Anwendung des

Beurteilungsschemas in der Praxis sollten die gewählten Klassifikationsregeln an einem grösseren Datensatz überprüft und allenfalls angepasst werden.

Ein wesentlicher Unterschied zwischen Expertenurteil und neuem Auswertungsmodell liegt in der *Datenbasis, auf der die beiden Einschätzungen basieren*. Datengrundlage des Expertenurteils ist die szenariospezifische Auswertung, so wie sie auch in der Praxis angewendet wird, das heisst

- die *Graphik* über den Suchweg und die Suchstrategie sowie
- der umfangreiche *Protokollbogen*, bestehend aus verschiedenen Angaben und Kennzahlen zu ausgewählten Aspekten des Problemlöseverhaltens (Steuerung, Kommunikation, Informationssuche), den Inhalten aller vom Probanden verfassten Texteinheiten (Protokolle, Meldungen) sowie den 13 Leistungskennwerten.

Die Beurteilung basiert damit auf der Auszählung oder reinen Auflistung *ausgewählter Aktionen* (z.B. Anzahl Sonareinsätze, Lesezeit Hilfesystem, Zeitpunkt für ersten Einsatz der Steuerautomatik) sowie *ausgewählter Handlungssequenzen* (z.B. Anteil optimal versetzter Sonareinsätze, Auflistung des zeitlichen Ablaufs der Kommunikation). Daneben liefert die Auswertung auch Angaben über den *Gesamtprozess* (Graphik) und über das *mentale Modell* des Probanden, über sein Denken, seine Absichten usw. (verfasste Protokolle und Meldungen). All diese verschiedenen Aspekte müssen interpretiert und irgendwie zu einem Urteil hinsichtlich Informationseröffnendem und Adäquatem Handeln integriert werden.

Datengrundlage des neuen Auswertungsmodells sind die *Verhaltensprotokolle*, das heisst die zeitsynchrone Auflistung *aller Handlungsschritte* des Probanden und des *zugehörigen Handlungskontextes* (Systemzustand, Systemmeldungen), sowie – analog zum Expertenurteil – die Inhalte aller vom Probanden verfassten *Texteinheiten*. Jedes Mikroverhalten wird im aktuellen Handlungskontext interpretiert und hinsichtlich Informationseröffnendem und Adäquatem Handeln bewertet. Im Vordergrund dieser Bewertungen steht nicht der Gesamtprozess der Problemlösung, sondern der „*Mikroprozess*“, das was vor und nach einer bestimmten Aktion passiert¹⁰. Das führt

¹⁰ Der Gesamtprozess der Problembearbeitung spielt beim neuen Auswertungsmodell erst nach der Mikroverhaltensbeschreibung eine Rolle, indem dargestellt (und durch die Prozessprofile bewertet) wird, wie sich der Proband während der Problembearbeitung in Bezug auf das Informationseröffnende und Adäquate Handeln verhält.

dazu, dass bestimmte Aktionen, die grundsätzlich für die Problemlösung sinnvoll und relevant sind (z.B. Sonar- und Videoeinsätze mit Fund im Videosichtbereich) je nach Kontext auch negativ bewertet werden können. Auch der Inhalt von Texteinheiten kann unter Berücksichtigung der vorausgegangenen oder folgenden Aktionen eine andere Bewertung erhalten (z.B. kann die Meldung „Habe bis jetzt nichts gefunden“, je nachdem ob davor Sonaraufnahmen mit oder ohne Meeresgrund erfolgten, anders bewertet werden). Bis zu einem gewissen Grade kann natürlich auch der Experte (mit etwas Aufwand) Kontextanalysen vornehmen, allerdings nicht in Bezug auf einen zeitlich gleich engen Handlungskontext, nicht in Bezug auf den selben „Mikrokontext“, da der Protokollbogen und die Graphik dies nicht zur Verfügung stellen.

Experte und Auswertungsmodell betrachten das Problemlöseverhalten aus unterschiedlichen Perspektiven, der Blickwinkel (im Sinne des Scheinwerferkegels) ist nicht der gleiche. Von daher ist es nicht erstaunlich, dass die Beurteilungen nicht exakt übereinstimmen.

Beim Vergleich von Expertenurteil und neuem Auswertungsmodell darf auch eine gewisse *Willkür in verschiedenen Schritten der Urteilsbildung* auf beiden Seiten nicht ausgeschlossen werden. Beim Experten ist unklar, welche der vielen verschiedenen Angaben über das Problemlöseverhalten mehr fokussiert wurden als andere (z.B. falsche Sonar- und Videoeinsätze versus Wiederholungen beim Lesen) und wie die berücksichtigten Aspekte integriert wurden. Analog zum Vorgehen in der Praxis wurden *weder Interpretations- noch Integrationsregeln* definiert. Um mehr über diese Willkür und mögliche Urteilsverzerrungen zu erfahren, hätten mehrere Urteiler einbezogen werden müssen (Interraterreliabilität). Aufgrund mangelnder finanzieller und damit auch personeller Ressourcen war dies leider nicht möglich.

Aber auch wenn, wie beim neuen Auswertungsmodell, Regeln zur Standardisierung der Dateninterpretation und -integration definiert werden, ist die Einschätzung der Problemlöseleistung nicht frei von Willkür. Diese liegt genau in der Definition dieser Regeln, das heisst in den gewählten *Interpretationsregeln der Mikroverhaltensbeschreibung* und den *Klassifikationsregeln zur Definition der Prozessprofile*. Letzteres wurde weiter oben bereits diskutiert (s. „Grenzfälle“ und „Stichprobenabhängigkeit“). Aus-

serdem liess sich die vorgenommene Klassifizierung mehrheitlich gut durch die formalen Klassifizierungsmethoden bestätigen, was gegen eine zu grosse Willkür der definierten Klassifikationsregeln und Prozessprofile spricht. Inwiefern aber eine *Überarbeitung der Interpretationsregeln*, das heisst andere Codierungen des Mikroverhaltens zu anderen Prozessprofilen und damit schliesslich zu einer anderen (besseren?) Einschätzung der Problemlöseleistung führt, muss in weiteren Arbeiten überprüft werden. Im Folgenden werden einige kritische Punkte der Mikroverhaltensbeschreibung diskutiert.

Überlegungen zur Mikroverhaltensbeschreibung

In der jetzigen Version wurde beispielsweise der „Erfolg“ von Sonar- und Videoeinsätzen unabhängig von ihrem Ort beurteilt. Einsätze in Gebieten mit einer Meerestiefe von mehr als 50 Metern wurden gleich beurteilt wie Einsätze in den beiden 50-Meter-Gebieten S/7 und R/9. Beim Informationseröffnenden Handeln wurde dies dadurch gerechtfertigt, dass ein Proband immer Informationen über den Verbleib des Flugzeuges erhält, auch wenn in Anbetracht der Überlebenschancen der Passagiere eine Suche in 50-Meter-Gebieten angezeigt wäre. Ähnliches gilt für das Adäquate Handeln: Auch Probanden, die in weiten Tiefen suchen, können durchaus im aktuellen Handlungskontext konsistent und angemessen und entsprechend der generellen Zielsetzung handeln. Die Notfallmeldung, die den Probanden über den Flugzeugabsturz informiert, enthält zwar den Vermerk über die Überlebenschancen bis 50 Meter Tiefe (und damit den Hinweis auf die sinnvollen Suchgebiete S/7 und R/9), der Proband wird jedoch aufgefordert, „sofort die Suche nach dem vermissten Flugzeug aufzunehmen“. Die generelle Zielsetzung kann damit „Finde das Flugzeug“ oder in einer spezifischeren Form „Finde möglichst rasch überlebende Passagiere“ lauten. Beides wurde im neuen Auswertungsmodell gleich bewertet, während bei der szenariospezifischen Auswertung und damit auch beim Expertenurteil die *Einschränkung der Suche auf die beiden sinnvollen Gebiete* stark im Vordergrund steht. In einer Überarbeitung der Mikroverhaltensbeschreibung könnte deshalb korrektes Such- und Prüfverhalten in nicht sinnvollen Gebieten beispielsweise mit +0.5 (versus +1 bei sinnvollen Gebieten) bewertet werden.

Zu überdenken ist auch die Bewertung der *Videoeinsätze*: Positiv bewertet werden nur *korrekte* Einsätze (Kriterien: < 4 Meter über Grund, davor Sonareinsatz mit Fund im Videosichtbereich und dazwischen keine Ortsveränderung, korrekte Ausrichtung von Nautilus). Alle Einsätze, die eines der Kriterien nicht erfüllen, werden negativ bewertet (Info = 0, Ad = -1). Der Auszählung von „Videoeinsätzen mit Metall im Videosichtbereich“ bei der szenariospezifischen Auswertung liegen hingegen nicht die gleich strengen Kriterien zugrunde. Auch hier könnte in der Mikroverhaltensbeschreibung eine differenziertere Bewertung aussagekräftiger sein.

Ebenfalls weiter unterscheiden liesse sich die Bewertung von *Sonareinsätzen* mit Fund im Scanbereich und Sonareinsätzen mit Fund im Videosichtbereich, die in der jetzigen Version beide gleich bewertet werden. Im Zusammenhang mit einer Differenzierung der Bewertung bei bestimmten Aktionen muss auch eine Überarbeitung der *Bewertungsskala* (Info: +1, +0.5, 0; Ad: +1, 0, -1) in Betracht gezogen werden.

Weitere mögliche Änderungen betreffen die Bewertung von *Leseaktionen*, und dabei insbesondere die Frage, wann ein Proband eine Informationsseite (der Datenbank Nauti-Info, des Hilfesystems oder von eingehenden Meldungen) zum ersten Mal gelesen und wann er sie wiederholt gelesen hat. Für jede Informationsseite wurde aufgrund der Textlänge und des Informationsgehalts eine so genannte *minimale Lesedauer* festgelegt, das heisst eine Zeitdauer, während der die Seite geöffnet sein muss, damit sie als „gelesen“ oder bei erneutem Aufrufen als „wiederholt gelesen“ gilt (2, 5 oder 10 Sekunden). Ob die Probanden durch die gewählten Kriterien zu gut oder zu streng beurteilt werden, können die vorliegenden Daten nicht beantworten. Hierfür könnte es hilfreich sein zu prüfen, wie lange die einzelnen Seiten beim ersten oder wiederholten Aufrufen im Durchschnitt über alle 230 Probanden geöffnet sind. In dieser Arbeit wurde auf eine solche Auswertung verzichtet. Die Probanden verbringen in Relation zur Gesamtspielzeit wenig Zeit mit Lesen ($M_d = 3\%$ für die Datenbank Nauti-Info und 13% für das Hilfesystem), und zudem spielt das Leseverhalten auch bei der szenariospezifischen Auswertung im Vergleich zum Such- und Prüfverhalten eine untergeordnete Rolle.

Im Hinblick auf die Anwendung der neuen Auswertung in der diagnostischen Praxis muss auch überprüft werden, wie gut sich die Kriterien für die Bewertung der vom Probanden verfassten *Texteinheiten* programmieren lassen. Zumindest für die formalen Bewertungskriterien sollte dies einfach umzusetzen sein. Auch bei den inhaltlichen Bewertungskriterien sollte die Beurteilung der formalen Korrektheit und der Bedeutung im aktuellen Handlungskontext möglich sein. Die Beurteilung der inhaltlichen Verständlichkeit hingegen dürfte schwieriger zu realisieren sein.

Weiterführende Überlegungen zum Einsatz in Praxis und Forschung

Grundsätzlich können natürlich alle Bewertungskriterien der Mikroverhaltensbeschreibung hinterfragt und kritisiert werden. Wie an einigen Beispielen veranschaulicht, lassen sich einzelne Regeln so anpassen, dass eine höhere Übereinstimmung zwischen neuem Auswertungsmodell und szenariospezifischer Auswertung beziehungsweise Expertenurteil resultieren könnte. Aber: Da keine Aussenkriterien vorliegen, kann abschliessend nicht geklärt werden, *welche Beurteilung nun die richtige oder bessere ist*. Es ist zu bedauern, dass das Unternehmen, das ATLANTIS während mehrerer Jahre im Rahmen von Assessment Centern zum Zwecke der Personalauswahl und -entwicklung (an über 200 Probanden) einsetzte, keine zusätzlichen Daten aus diesen Erhebungen (z.B. Resultate anderer Leistungstests) oder Angaben über Berufserfolgskriterien zur Verfügung stellt. Es wäre interessant zu wissen, welche Beurteilung mehr Varianz in solchen Aussenkriterien aufklärt, und ob das neue Auswertungsmodell im Vergleich zur szenariospezifischen Auswertung *zusätzliche* Varianzanteile aufzuklären vermag. Wenn auch für die szenariospezifische Auswertung bis heute ebenfalls keine Validitätsangaben vorliegen, so spricht der mehrjährige Einsatz derselben zur Auswahl von Führungskräften in einem grossen, renommierten Dienstleistungsunternehmen doch zumindest ansatzweise für ihre Brauchbarkeit und Zuverlässigkeit. Der Einsatz des neuen Auswertungsmodells in der diagnostischen Praxis müsste daher mit dem Nachweis der inkrementellen Validität begründet werden.

Daneben würde auch eine hohe (bzw. im Vergleich zur bestehenden Auswertung höhere) *Akzeptanz auf Seiten der Diagnostiker* die Einführung der neuen Auswertung

in der Praxis legitimieren. Das neue Auswertungsmodell wurde mit dem Ziel konzipiert, die praktische Relevanz von Problemlöseszenarien zu erhöhen, das heisst die Anwendung solcher Instrumente in der Praxis zu erleichtern, indem Datenauswertung, -interpretation und -integration standardisiert werden und ein plausibles, ohne Expertenwissen und ohne viel Zeitaufwand anwendbares Auswertungsschema vorliegt. Dass die Einschätzungen der Probandenleistung durch das neue Auswertungsschema in die richtige Richtung gehen, zeigen die Befunde dieser Arbeit. Die mehrheitlich gute Übereinstimmung mit dem Expertenurteil spricht zudem dafür, dass sich die aufwändige Dateninterpretation durch den Diagnostiker ersetzen lässt. Natürlich könnte man nun einwenden, dass sich auch diese standardisieren und automatisieren liesse. Diesbezüglich wurden jedoch noch keine Ideen vorgelegt, und aus den Angaben des Experten über seine Urteilsbildung (vgl. Kapitel 3.3) lässt sich nicht entnehmen, ob Regeln zur Standardisierung der Dateninterpretation und -integration überhaupt operationalisierbar wären.

Inwiefern nun aber das neue Auswertungsmodell tatsächlich auf Anwenderseiten begrüsst wird, konnte nicht geprüft werden. Im Verlaufe des Dissertationsprojekts wurde der Einsatz von ATLANTIS in dem besagten Unternehmen eingestellt, weil „die Software veraltet sei hinsichtlich Layout und Aufwand für die Auswertung“¹¹. Eine Neuprogrammierung des Instruments ist zwar in Vorbereitung, es hätte aber den zeitlichen Rahmen der vorliegenden Dissertation gesprengt, das entwickelte Auswertungsmodell auf die neue Softwareversion anzuwenden und hinsichtlich Akzeptanz auf Anwenderseiten zu prüfen.

Die Aussage des Unternehmens, ATLANTIS sei im Hinblick auf die Auswertung zu aufwändig, spricht allerdings für einen künftigen Einsatz der neuen Auswertung in der Praxis. Statt der umfangreichen, mehrseitigen Datenausgabe, bestehend aus einer Vielzahl von Kennzahlen und Kennwerten erhält der Diagnostiker bei der neuen Auswertung nur zwei Resultate: a) die Zuordnung eines Probanden zu einem bestimmten Problemlösetypen (einem typischen Bearbeitungsmuster), der eine Bewertung der Problemlösegröße beinhaltet, und b) die Darstellung der individuellen

¹¹ Persönliche Mitteilung vom 11.1.2006 der Leiterin Personalauswahl im besagten Unternehmen.

Zeitreihen, das heisst der zeitlichen Entwicklung des Informationseröffnenden und Adäquaten Handelns, welche genauere Aussagen über den Bearbeitungsprozess ermöglichen. Die Angaben aus der szenariospezifischen Auswertung könnten nun verwendet werden, um das Zustandekommen dieser allgemeinen Einschätzung der Problemlöseleistung zu erklären. Wie bereits im Theorieteil erläutert, kann die Berücksichtigung von Verhaltensdaten und Verhaltensskalen wichtige Informationen für das diagnostische Abschlussfeedback oder für Personalentwicklungsmassnahmen liefern (vgl. auch U. Funke, 1995; Wagener & Wittmann, 2002). Das bedeutet: Die neue Auswertung soll die bestehende, szenariospezifische Auswertung nicht ersetzen, sondern vielmehr *ergänzen*. Der Vorteil der neuen Auswertung liegt in der Verbesserung der Interpretationsobjektivität und der Ökonomie, der Vorteil der bestehenden Auswertung in der Fülle an Informationen über unterschiedliche Verhaltensaspekte der Problembearbeitung.

Die bisherigen Ausführungen betrafen den praktischen Einsatz des neuen Auswertungsmodells. Aus der vorliegenden Arbeit können aber auch einige Konsequenzen für die Grundlagenforschung abgeleitet werden. Das neue Auswertungsmodell erfasst zwei *abstrakte, vom Inhalt des Szenarios unabhängige Konstrukte*, die in der Literatur als grundlegend für das Lösen komplexer Probleme aufgeführt werden. Die beiden Konstrukte erfassen, wie effizient und konsistent der Proband problemrelevante Informationen sucht und wie gut er die (potentiell) aufgenommenen Informationen umsetzen und nutzen kann. Damit wird etwas Ähnliches abgebildet wie bei den Ansätzen zur Erfassung von Wissenserwerb und Wissensanwendung. Die dafür vorgeschlagenen Methoden berechnen jedoch vorwiegend globale Masse, die keine Aussage über den *Prozess der Bearbeitung* erlauben (eine Ausnahme findet sich bei Wirth, 2004; Wirth & Funke, 2005). Ausserdem beschränkt sich die Erfassung von Wissenserwerb und Wissensanwendung auf artifizielle, formal orientierte Szenarien (J. Funke, 2003). Das vorliegende Auswertungsmodell hingegen ist prozessorientiert und beschränkt sich nicht auf bestimmte (inhaltlich oder formal orientierte) Szenarien. Die einzige Voraussetzung betrifft die Datenspeicherung: Das Problemlöseverhalten muss zeitsynchron gespeichert werden, wobei für jeden Handlungsschritt auch der zugehörige Handlungskontext (Systemzustand) gespei-

chert werden muss. Bei der Weiterentwicklung von Szenarien sollte dies berücksichtigt werden. Denn gerade in der Möglichkeit zur Prozessdiagnostik liegt der Vorteil der Szenarien gegenüber herkömmlichen Leistungstests.

Durch die Einschätzung der Problemlöseleistung durch szenario-unspezifische (Prozess-) Masse rückt das *komplexe Problemlösen* näher an das Intelligenzkonstrukt, wo ebenfalls versucht wird, *abstrakte* Dimensionen zu erfassen. Die beiden Konstrukte liegen damit – im Sinne von Wittmann (1987) und Wittmann und Matt (1986) – auf dem (annähernd) *gleichen Generalitätsniveau*, was auf szenariospezifische Masse weniger zutrifft. Deshalb müssten die Korrelationen zwischen Szenarien und Intelligenztests höher ausfallen als in bisherigen Untersuchungen mit szenariospezifischen Massen berichtet wurde. Auch ist zu erwarten, dass Untersuchungen über die eignungsdiagnostischen Gütekriterien von Szenarien weniger heterogene Befunde liefern würden, wenn in allen Untersuchungen die Problemlöseleistung durch die gleichen szenario-unspezifischen Prozessmasse eingeschätzt wird. Inwiefern solche Spekulationen berechtigt sind, wird die weitere Forschung zum „komplexen Problemlösen“ (hoffentlich) zeigen.

6 Literaturverzeichnis

- Arbeitskreis Assessment Center e. V. (1992). *Standards der Assessment Center Technik* [Online]. Available: <http://www.arbeitskreis-ac.de>
- Arbeitskreis Assessment Center e. V. (2004). *Standards der Assessment Center Technik* [Online]. Available: <http://www.arbeitskreis-ac.de>
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler* (6. Aufl.). Heidelberg: Springer.
- Bortz, J. & Döring, N. (2002). *Forschungsmethoden und Evaluation für Human- und Sozialwissenschaftler* (3. überarb. Aufl.). Berlin: Springer.
- Bortz, J. & Lienert, G. A. (2003). *Kurzgefasste Statistik für die klinische Forschung: Leitfaden für die verteilungsfreie Analyse kleiner Stichproben* (2. aktual. und bearb. Aufl.). Berlin: Springer.
- Bortz, J., Lienert, G. A. & Boehnke, K. (2000). *Verteilungsfreie Methoden in der Biostatistik* (2. Aufl.). Berlin: Springer.
- Breuer, K., Molkenthin, R. & Tennyson, R. D. (2006). Role of simulation in web-based learning. In H. F. O'Neil & R. S. Perez (Eds.), *Web-based learning. Theory, research, and practice* (pp. 307-326). Mahwah, NJ: Lawrence Erlbaum Associates.
- Buchner, A. (1999). Komplexes Problemlösen vor dem Hintergrund der Theorie finiter Automaten. *Psychologische Rundschau*, 50 (4), 206-212.
- Buchner, A. & Funke, J. (1993). Finite-state automata: Dynamic task environments in problem-solving research. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 46A (1), 83-118.
- Cohen, J. (1968). Weighted Kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin*, 70 (4), 213-220.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- De Jong, T. & van Joolingen, W. R. (1998). Scientific discovery learning with computersimulations of conceptual domains. *Review of Educational Research*, 68, 179-201.
- Dörner, D. (1976). *Problemlösen als Informationsverarbeitung*. Stuttgart: Kohlhammer.
- Dörner, D. (1986). Diagnostik der operativen Intelligenz. *Diagnostica*, 32 (4), 290-308.
- Dörner, D. (1989). *Die Logik des Misslingens. Strategisches Denken in komplexen Situationen*. Reinbek: Rowohlt.
- Dörner, D. (1999). *Bauplan für eine Seele*. Reinbek: Rowohlt.
- Dörner, D., Bartl, C., Detje, F., Gerdes, J., Halcour, D., Schaub, H. & Starker, U. (2002). *Die Mechanik des Seelenwagens: eine neuronale Theorie der Handlungsregulation*. Bern: Huber.
- Dörner, D., Kreuzig, H. W., Reither, F. & Stäudel, T. (1983). *Lohhausen. Vom Umgang mit Unbestimmtheit und Komplexität*. Bern: Hans Huber.
- Elshout, J. J. (1987). Problem solving and education. In R. Parmentier, E. de Corte & H. Lodewijks (Eds.), *Learning and instruction: European research in an international context, Vol. 1* (pp. 259-273). Elmsford, NY: Pergamon.
- Frensch, P. A. & Funke, J. (1995). *Complex problem solving: The European perspective*. Hillsdale, NJ: Lawrence Erlbaum Associates.

- Funke, J. (1983). Einige Bemerkungen zu Problemen der Problemlöseforschung oder: Ist Testintelligenz doch ein Prädiktor? *Diagnostica*, 29 (4), 283-302.
- Funke, J. (1993). Microworlds based on linear equation systems: A new approach to complex problem solving and experimental results. In G. Strube (Ed.), *The cognitive psychology of knowledge* (pp. 313-330). Amsterdam: Elsevier Science Publishers.
- Funke, J. (1998). Computer-based testing and training with scenarios from complex problem-solving research: Advantages and disadvantages. *International Journal of Selection and Assessment*, 6 (2), 90-96.
- Funke, J. (1999). Komplexes Problemlösen – Ein Blick zurück und ein Blick nach vorne. *Psychologische Rundschau*, 50 (4), 194-197.
- Funke, J. (2001). Neue Verfahren zur Erfassung intelligenten Umgangs mit komplexen und dynamischen Anforderungen. In E. Stern & J. Guthke (Hrsg.), *Perspektiven der Intelligenzforschung. Ein Lehrbuch für Fortgeschrittene* (S. 89-107). Lengerich: Pabst Science Publishers.
- Funke, J. (2003). *Problemlösendes Denken*. Stuttgart: Kohlhammer.
- Funke, J. (2006a). Denken und Problemlösen, *Enzyklopädie der Psychologie, Themenbereich C, Serie II, Band 8*. Göttingen: Hogrefe.
- Funke, J. (2006b). Komplexes Problemlösen. In J. Funke (Hrsg.), *Denken und Problemlösen, Enzyklopädie der Psychologie, Themenbereich C, Serie II, Band 8* (S. 375-446). Göttingen: Hogrefe.
- Funke, J. & Frensch, P. A. (2005). *Handbuch der allgemeinen Psychologie – Kognition*. Göttingen: Hogrefe.
- Funke, U. (1993). Computergestützte Eignungsdiagnostik mit komplexen dynamischen Szenarios. *Zeitschrift für Arbeits- und Organisationspsychologie*, 37, 109-118.
- Funke, U. (1995). Szenarien in der Eignungsdiagnostik und im Personaltraining. In B. Strauss & M. Kleinmann (Hrsg.), *Computersimulierte Szenarien in der Personalarbeit* (S. 145-216). Göttingen: Verlag für Angewandte Psychologie.
- Geilhardt, T. & Mühlbradt, T. (1995). *Planspiele im Personal- und Organisationsmanagement*. Göttingen: Verlag für Angewandte Psychologie.
- Gonzalez, C., Vanyukov, P. & Martin, M.-K. (2005). The use of microworlds to study dynamic decision making. *Computers in Human Behavior*, 21, 273-286.
- Gray, W. D. (2002). Simulated task environments: The role of high-fidelity simulations, scaled worlds, synthetic environments, and laboratory tasks in basic and applied cognitive research. *Cognitive Science Quarterly*, 2, 205-227.
- Hartung, S. & Schneider, I. (1995). Entwicklung und Anwendung computersimulierter Szenarien. In B. Strauss & M. Kleinmann (Hrsg.), *Computersimulierte Szenarien in der Personalarbeit* (S. 219-236). Göttingen: Verlag für Angewandte Psychologie.
- Hasselmann, D. (1995). Die Konstruktion computersimulierter Szenarien für die Personalarbeit. In B. Strauss & M. Kleinmann (Hrsg.), *Computersimulierte Szenarien in der Personalarbeit* (S. 237-259). Göttingen: Verlag für Angewandte Psychologie.
- Hirsig, R. & de With, A. E. (1991). Interaktives Computerspiel – Neues Instrument für die Psychodiagnostik. *Psychoscope*, 12 (10), 12-14.

- Hirsig, R. & de With, A. E. (1992). STRATOS-00: Ein PC-gestütztes Erhebungsinstrument zum „integralen Problem-Management“. In U. Imoberdorf, R. Käser & R. Zihlmann (Hrsg.), *Psychodiagnostik heute. Beiträge aus Theorie und Praxis* (S. 105-123). Stuttgart: Hirzel.
- Hirsig, R. & de With, A. E. (1996). *Computer Simulation ATLANTIS*. Zürich: Computer anwendende Techniken (cat).
- Hirsig, R. & de With, A. E. (1998). PC-gestützte Simulationen für die Diagnostik: Fallbeispiel ATLANTIS. In U. Imoberdorf, R. Käser & R. Zihlmann (Hrsg.), *Psychodiagnostik von Individuen, Gruppen und Organisationen* (S. 141-160). Stuttgart: Hirzel.
- Hirsig, R. & de With, A. E. (2001). Testkurzdarstellung ATLANTIS. In W. Sarges & H. Wottawa (Hrsg.), *Handbuch wirtschaftspsychologischer Testverfahren* (S. 67-70). Lengerich: Pabst.
- Höft, S. & Funke, U. (2006). Simulationsorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (2. Aufl., S. 145-188). Göttingen: Hogrefe.
- Hornke, L.-F. & Kersting, M. (2006). Optimizing quality in the use of web-based and computer-based testing for personnel selection. In D. Bartram & R. K. Hambleton (Eds.), *Computer-based testing and the internet. Issues and advances* (pp. 149-162). Chichester: Wiley.
- Hossiep, R. (2001). Psychologische Tests – die vernachlässigte Dimension in Assessment Centern. In W. Sarges (Hrsg.), *Weiterentwicklungen der Assessment Center-Methode* (S. 53-68). Göttingen: Verlag für Angewandte Psychologie.
- Jansson, A. (1994). Pathologies in dynamic decision making: Consequences or precursors of failure? *Sprache & Kognition*, 13, 160-173.
- Jonassen, D. H. (2000). Toward a design theory of problem solving. *Educational Technology Research & Development*, 48 (4), 63-85.
- Kallus, W. K. & Janke, W. (1995). Klassenzuordnung. In R. S. Jäger & F. Petermann (Hrsg.), *Psychologische Diagnostik: ein Lehrbuch* (S. 170-186). München: Weinheim.
- Kersting, M. (1998). Differentielle Aspekte der sozialen Akzeptanz von Intelligenztests und Problemlöseszenarien als Personalauswahlverfahren. *Zeitschrift für Arbeits- und Organisationspsychologie*, 42 (2), 61-75.
- Kersting, M. (1999a). Computergestützte Problemlöseszenarien in der (Eignungs-) Diagnostik – (Charakteristische?) Defizite beim Forschungs-Praxis-Transfer. In W. Hacker & M. Rinck (Hrsg.), *Bericht über den 41. Kongress der Deutschen Gesellschaft für Psychologie in Dresden 1998. Schwerpunktthema „Zukunft gestalten“* (S. 357-365). Lengerich: Pabst.
- Kersting, M. (1999b). *Diagnostik und Personalauswahl mit computergestützten Problemlöseszenarien? Zur Kriteriumsvalidität von Problemlöseszenarien und Intelligenztests*. Göttingen: Hogrefe.
- Kersting, M. (2001). Zur Konstrukt- und Kriteriumsvalidität von Problemlöseszenarien anhand der Vorhersage von Vorgesetztenurteilen über die berufliche Bewährung. *Diagnostica*, 47 (2), 67-76.
- Kersting, M. & Kluge, A. (2001). Innovative eignungsdiagnostische Instrumente und Trainingstools – warum sie häufig scheitern. *Wirtschaftspsychologie*, 2, 77-86.
- Kleinmann, M. (2003). *Assessment-Center*. Göttingen: Hogrefe.

- Kleinmann, M. & Strauss, B. (1998). Validity and application of computer-simulated scenarios in personnel assessment. *International Journal of Selection and Assessment*, 6 (2), 97-106.
- Kluwe, R. H., Misiak, C. & Haider, H. (1991). The control of complex systems and performance in intelligence tests. In H. Rowe (Ed.), *Intelligence: reconceptualization and measurement* (pp. 227-244). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kluwe, R. H., Schilde, A., Fischer, C. & Oellerer, N. (1991). Problemlöseleistungen beim Umgang mit komplexen Systemen und Intelligenz. *Diagnostica*, 37 (4), 291-313.
- Kröner, S. (2001). *Intelligenzdiagnostik per Computersimulation*. Münster: Waxmann.
- Kröner, S., Plass, J.-L. & Leutner, D. (2005). Intelligence assessment with computer simulations. *Intelligence*, 33, 347-368.
- Leutner, D. (1995). Computerunterstützte Planspiele als Instrument der Personalentwicklung. In T. Geilhardt & T. Mühlbradt (Hrsg.), *Planspiele im Personal- und Organisationsmanagement* (S. 105-116). Göttingen: Verlag für Angewandte Psychologie.
- Leutner, D. (2002). The fuzzy relationship of intelligence and problem solving in computer simulations. *Computers in Human Behavior*, 18, 685-697.
- Milligan, G. W. (1981). A review of Monte Carlo tests of cluster analysis. *Multivariate Behavioral Research*, 16 (3), 379-407.
- Müller, B. & Funke, J. (1995). Das Paradigma „Komplexes Problemlösen“. In B. Strauss & M. Kleinmann (Hrsg.), *Computersimulierte Szenarien in der Personalarbeit* (S. 57-102). Göttingen: Verlag für Angewandte Psychologie.
- Müller, H. (1993). *Komplexes Problemlösen: Reliabilität und Wissen*. Bonn: Holos.
- Naef, B. (1996). *Komplexe Problemlösesimulationen und ihr Einsatz als diagnostisches Instrument*. Unveröff. Lizenziatsarbeit, Universität Zürich, Psychologisches Institut, Psychologische Methodenlehre.
- Naef, B. (2001a). Kontextunabhängige Verhaltensstrategien in zwei Simulationen zur Einschätzung der Problemlöseleistung. In U. Gerhard (Hrsg.), *Psychologie und Lebensqualität* (S. 197-199). Freiburg: Universitätsverlag.
- Naef, B. (2001b). *Wissenschaftliche Theorie und Anforderungen der Praxis in der Personalselektion. Ein Versuch zur Synthese* (Dissertation, Universität Zürich). Zürich: Zentralstelle der Studentenschaft.
- Obermann, C. (2002). *Assessment-Center: Entwicklung, Durchführung, Trends*. Wiesbaden: Gabler.
- Putz-Osterloh, W. (1983). Über Determinanten komplexer Problemlöseleistungen und Möglichkeiten zu ihrer Erfassung. *Sprache & Kognition*, 2 (2), 100-116.
- Putz-Osterloh, W. (1993). Complex problem solving as a diagnostic tool. In H. Schuler, J. L. Farr & M. Smith (Eds.), *Personnel selection and assessment: Individual and organizational perspectives* (pp. 289-301). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Raaheim, K. (1988). Intelligence and task novelty. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence, Vol. 4* (pp. 73-97). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Rigas, G., Carling, E. & Brehmer, B. (2002). Reliability and validity of performance measures in microworlds. *Intelligence*, 30, 463-480.

- Sarges, W. (2000). *Management-Diagnostik* (3. unveränd. Aufl.). Göttingen: Hogrefe.
- Sarges, W. (2001a). Einleitung des Herausgebers: Die Assessment Center-Methode – Herkunft, Kritik und Weiterentwicklungen. In W. Sarges (Hrsg.), *Weiterentwicklungen der Assessment Center-Methode* (S. VII-XXXII). Göttingen: Hogrefe.
- Sarges, W. (2001b). *Weiterentwicklungen der Assessment Center-Methode*. Göttingen: Hogrefe.
- Schaub, H. (1990). Die Situationsspezifität des Problemlöseverhaltens. *Zeitschrift für Psychologie*, 198, 83-96.
- Schaub, H. (2006). Störungen und Fehler beim Denken und Problemlösen. In J. Funke (Hrsg.), *Denken und Problemlösen, Enzyklopädie der Psychologie, Themenbereich C, Serie II, Band 8* (S. 447-482). Göttingen: Hogrefe.
- Schmidt, F. L. & Hunter, J. E. (1998a). Messbare Personmerkmale: Stabilität, Variabilität und Validität zur Vorhersage zukünftiger Berufsleistung und berufsbezogenen Lernens. In M. Kleinmann & B. Strauss (Hrsg.), *Potentialfeststellung und Personalentwicklung* (S. 15-43). Göttingen: Hogrefe.
- Schmidt, F. L. & Hunter, J. E. (1998b). The validity and utility of selection methods in personnel psychology. Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.
- Schmuck, P. (1992). Zum Zusammenhang zwischen der Effizienz exekutiver Kontrolle und dem mehrfachen Lösen eines komplexen Problems. *Sprache & Kognition*, 11 (4), 193-207.
- Schmuck, P. & Strohschneider, S. (1995). Exekutive Kontrolle und Verhaltensstabilität beim Bearbeiten eines komplexen Problems: Eine Replikation. *Diagnostica*, 41 (2), 150-171.
- Schöning, H. (1995). Computersimulierte Szenarien in einem internationalen Konzern (Beiersdorf AG). In B. Strauss & M. Kleinmann (Hrsg.), *Computersimulierte Szenarien in der Personalarbeit* (S. 261-271). Göttingen: Verlag für Angewandte Psychologie.
- Schoppek, W. (1991). Spiel und Wirklichkeit – Reliabilität und Validität von Verhaltensmustern in komplexen Situationen. *Sprache & Kognition*, 10 (1), 15-27.
- Schoppek, W. (1996). *Kompetenz, Kontrollmeinung und komplexe Probleme. Zur Vorhersage individueller Unterschiede bei der Systemsteuerung*. Bonn: Holos.
- Schoppek, W. (2002). Examples, rules, and strategies in the control of dynamic systems. *Cognitive Science Quarterly*, 2, 63-92.
- Schoppek, W. & Putz-Osterloh, W. (2003). Individuelle Unterschiede und die Bearbeitung komplexer Probleme. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 24 (3), 163-173.
- Schreiber, L. (1995). Der Einsatz eines computersimulierten Szenarios im Assessment-Center bei MERCK. In B. Strauss & M. Kleinmann (Hrsg.), *Computersimulierte Szenarien in der Personalarbeit* (S. 273-284). Göttingen: Verlag für Angewandte Psychologie.
- Schuler, H. & Funke, U. (1993). Diagnose beruflicher Eignung und Leistung. In H. Schuler (Hrsg.), *Lehrbuch Organisationspsychologie* (S. 235-283). Bern: Huber.
- Schuler, H. & Höft, S. (2004a). Berufseignungsdiagnostik und Personalauswahl. In H. Schuler (Hrsg.), *Organisationspsychologie – Grundlagen und Personalpsychologie, Enzyklopädie der Psychologie, Themenbereich D, Serie III, Band 3* (S. 439-532). Göttingen: Hogrefe.
- Schuler, H. & Höft, S. (2004b). Diagnose beruflicher Eignung und Leistung. In H. Schuler (Hrsg.), *Lehrbuch Organisationspsychologie* (S. 289-343). Bern: Huber.

- Schuler, H. & Höft, S. (2006). Konstruktorientierte Verfahren der Personalauswahl. In H. Schuler (Hrsg.), *Lehrbuch der Personalpsychologie* (S. 101-144). Göttingen: Hogrefe.
- Schweizer, K. (2006a). Intelligenz. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 2-15). Berlin: Springer.
- Schweizer, K. (2006b). Intelligenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 70-83). Berlin: Springer.
- Spering, M., Wagener, D. & Funke, J. (2005). The role of emotions in complex problem-solving. *Cognition and Emotion*, 19 (8), 1252-1261.
- Strauss, B. (2000). Die Messung der praktischen Intelligenz von Managern mit Hilfe computersimulierter Szenarien. In L. von Rosenstiel & T. Lang-von Wins (Hrsg.), *Perspektiven der Potentialbeurteilung* (S. 129-153). Göttingen: Hogrefe.
- Strauss, B. & Kleinmann, M. (1995). *Computersimulierte Szenarien in der Personalarbeit*. Göttingen: Verlag für Angewandte Psychologie.
- Strauss, B. & Kleinmann, M. (2001). Computersimulierte Szenarien im Assessment Center. In W. Sarges (Hrsg.), *Weiterentwicklungen der Assessment Center-Methode* (S. 69-86). Göttingen: Hogrefe.
- Streufer, S., Pogash, R. & Piasecki, M. (1988). Simulation-based assessment of managerial competence: Reliability and validity. *Personnel Psychology*, 41 (3), 537-557.
- Strohschneider, S. (1986). Zur Stabilität und Validität von Handeln in komplexen Realitätsbereichen. *Sprache & Kognition*, 1, 42-48.
- Strohschneider, S. (1990). *Wissenserwerb und Handlungsregulation*. Wiesbaden: Deutscher Universitäts-Verlag.
- Strohschneider, S. (1991). Problemlösen und Intelligenz: Über die Effekte der Konkretisierung komplexer Probleme. *Diagnostica*, 37 (4), 353-371.
- Strohschneider, S. (1992). *Intelligenz und Strategien beim Umgang mit einem komplexen Problem: Ein deutsch-deutscher Vergleich* (Memorandum Nr. 7). Universität Bamberg, Lehrstuhl Psychologie II.
- Strohschneider, S. (1994). Strategien beim Umgang mit einem komplexen Problem: Ein deutsch-deutscher Vergleich. *Zeitschrift für Arbeits- und Organisationspsychologie*, 38 (1), 34-40.
- Strohschneider, S. (1996). Strategien beim Umgang mit einem zieloffenen komplexen Problem. In S. Strohschneider (Hrsg.), *Denken in Deutschland* (S. 71-95). Bern: Huber.
- Strohschneider, S. (2001). Strategien des Handelns. Kulturvergleichende Untersuchungen zur Entwicklung strategischer Kompetenzen. In G. Franke (Hrsg.), *Komplexität und Kompetenz. Ausgewählte Fragen der Kompetenzforschung* (S. 327-355). Bielefeld: Bertelsmann.
- Strohschneider, S. (2006). Kulturelle Unterschiede beim Problemlösen. In J. Funke (Hrsg.), *Denken und Problemlösen, Enzyklopädie der Psychologie, Themenbereich C, Serie II, Band 8* (S. 549-618). Göttingen: Verlag für Psychologie.
- Süss, H.-M. (1996). *Intelligenz, Wissen und Problemlösen. Kognitive Voraussetzungen für erfolgreiches Handeln bei computersimulierten Problemen*. Göttingen: Hogrefe.
- Süss, H.-M. (1999). Intelligenz und komplexes Problemlösen. Perspektiven für eine Kooperation zwischen differentiell-psychometrischer und kognitionspsychologischer Forschung. *Psychologische Rundschau*, 50 (4), 220-228.

- Süss, H.-M. (2001). Die Rolle von Intelligenz und Wissen für erfolgreiches Handeln in komplexen Problemsituationen. In G. Franke (Hrsg.), *Komplexität und Kompetenz. Ausgewählte Fragen der Kompetenzforschung* (S. 249-275). Bielefeld: Bertelsmann.
- Süss, H.-M., Kersting, M. & Oberauer, K. (1991). Intelligenz und Wissen als Prädiktoren für Leistungen bei computersimulierten komplexen Problemen. *Diagnostica*, 37 (4), 334-352.
- Süss, H.-M., Kersting, M. & Oberauer, K. (1993). Zur Vorhersage von Steuerungsleistungen an computersimulierten Systemen durch Wissen und Intelligenz. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 14 (3), 189-203.
- Süss, H.-M., Oberauer, K. & Kersting, M. (1993). Intellektuelle Fähigkeiten und die Steuerung komplexer Systeme. *Sprache & Kognition*, 12 (2), 83-97.
- Swaak, J., de Jong, T. & van Joolingen, W. R. (2004). The effects of discovery learning and expository instruction on the acquisition of definitional and intuitive knowledge. *Journal of Computer Assisted Learning*, 20, 225-234.
- Wagener, D. (2001a). Personalauswahl und -entwicklung mit komplexen Szenarios. Qualifizierte und unzureichende Diagnostikinstrumente. *Wirtschaftspsychologie*, 2, 69-76.
- Wagener, D. (2001b). *Psychologische Diagnostik mit komplexen Szenarios: Taxonomie, Entwicklung, Evaluation*. Lengerich: Pabst.
- Wagener, D. & Wittmann, W. W. (2002). Personalarbeit mit dem komplexen Szenario FSYS: Validität und Potential von Verhaltensskalen. *Zeitschrift für Personalpsychologie*, 2, 80-93.
- Weber, H. & Westmeyer, H. (2001). Die Inflation der Intelligenzen. In E. Stern & J. Guthke (Hrsg.), *Perspektiven der Intelligenzforschung* (S. 251-266). Lengerich: Pabst Science Publishers.
- Weise, G. (1975). *Psychologische Leistungstests*. Göttingen: Hogrefe.
- Wirth, J. (2004). *Selbstregulation von Lernprozessen*. Münster: Waxmann.
- Wirth, J. & Funke, J. (2005). Dynamisches Problemlösen: Entwicklung und Evaluation eines neuen Messverfahrens zum Steuern komplexer Systeme. In E. Klieme, D. Leutner & J. Wirth (Hrsg.), *Problemlösekompetenz von Schülerinnen und Schülern. Diagnostische Ansätze, theoretische Grundlagen und empirische Befunde der deutschen PISA-2000-Studie* (S. 55-72). Wiesbaden: Verlag für Sozialwissenschaften.
- Wittmann, W. W. (1987). Grundlagen erfolgreicher Forschung in der Psychologie: Multimodale Diagnostik, Multiplismus, multivariate Reliabilitäts- und Validitätstheorie. *Diagnostica*, 33 (3), 209-226.
- Wittmann, W. W. & Matt, G. E. (1986). Aggregation und Symmetrie. Grundlagen einer multivariaten Reliabilitäts- und Validitätstheorie, dargestellt am Beispiel der differentiellen Validität des Berliner Intelligenzstrukturmodells. *Diagnostica*, 32 (4), 309-329.
- Wittmann, W. W. & Süss, H.-M. (1999). Investigating the paths between working memory, intelligence, knowledge, and complex problem-solving performances via brunswik symmetry. In P. L. Ackermann, P. C. Kyllonen & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and content* (pp. 77-108). Washington: American Psychological Association.
- Wottawa, H. & Hossiep, R. (1997). *Anwendungsfelder psychologischer Diagnostik*. Göttingen: Hogrefe.

7 Anhang

Anhang A:	Abbildungs- und Tabellenverzeichnis	162
Anhang B:	Ergänzungen zum Szenario ATLANTIS.....	164
Anhang C:	Ergänzungen zur Mikroverhaltensbeschreibung	174
Anhang D:	Ergänzungen zu den Ergebnissen.....	181
	Ergänzungen zu Kapitel 4.1	181
	Ergänzungen zu Kapitel 4.2.1	184
	Ergänzungen zu Kapitel 4.2.2	189
	Ergänzungen zu Kapitel 4.4.1	191
	Ergänzungen zu Kapitel 4.4.2 und Kapitel 4.4.3.....	192

Anhang A: Abbildungs- und Tabellenverzeichnis

Abbildungsverzeichnis

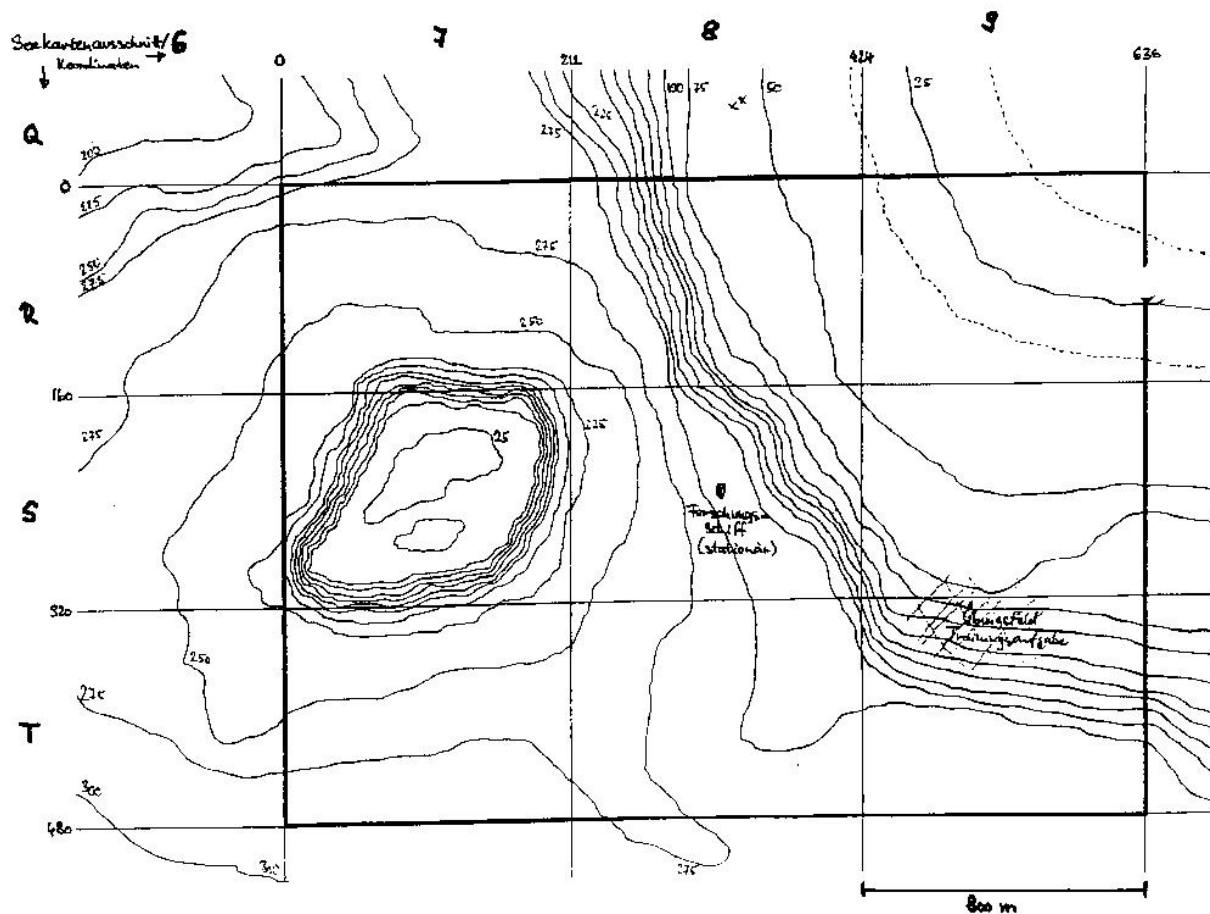
<i>Abbildung 1:</i>	Übersicht über das neue Auswertungsmodell	46
<i>Abbildung 2:</i>	Steuerungscomputer für das ferngesteuerte Mini-U-Boot Nautilus	53
<i>Abbildung 3:</i>	Graphische Darstellung des Suchwegs und der Suchstrategie bei der Bearbeitung von ATLANTIS	59
<i>Abbildung 4:</i>	Darstellung der zeitlichen Entwicklung des Informationseröffnenden Handelns und des Adäquaten Handelns	97
<i>Abbildung 5:</i>	Beispiele unterschiedlicher Problembearbeitungen.....	98
<i>Abbildung 6:</i>	Entwicklung des Informationseröffnenden und Adäquaten Handelns in der Teilstichprobe 1	102
<i>Abbildung 7:</i>	Verteilung der Spielzeit in der Teilstichprobe 1	103
<i>Abbildung 8:</i>	Dendrogramm der Clusteranalyse über die Ereignispunkte für das Informationseröffnende Handeln.....	113
<i>Abbildung 9:</i>	Ergebnis der MDS für das Informationseröffnende Handeln (Berechnung der Euklidischen Distanzen über Variante „Fixed Step“).....	116
<i>Abbildung 10:</i>	Dendrogramm der Clusteranalyse über die Ereignispunkte für das Adäquate Handeln.....	118
<i>Abbildung 11:</i>	Ergebnis der MDS für das Adäquate Handeln (Berechnung der Euklidischen Distanzen über Variante „Fixed Step“).....	120
<i>Abbildung 12:</i>	Verteilung der 13 Leistungskennwerte in der Teilstichprobe 1	128
<i>Abbildung 13:</i>	Liniendiagramm der Mediane in den szenariospezifischen Leistungskenn- werten für die Prozessprofile des Informationseröffnenden Handelns.....	131
<i>Abbildung 14:</i>	Liniendiagramm der Mediane in den szenariospezifischen Leistungskenn- werten für die Prozessprofile des Adäquaten Handelns	132
<i>Abbildung 15:</i>	Seekarte über das aktuelle Forschungsgebiet von Nautilus	164
<i>Abbildung 16:</i>	Beispiel eines Verhaltensprotokolls.....	174
<i>Abbildung 17:</i>	Verteilungen der 13 Leistungskennwerte in den untersuchten Stichproben	181
<i>Abbildung 18:</i>	Darstellung der Zeitreihen für die Teilstichprobe 1.....	184

Tabellenverzeichnis

<i>Tabelle 1:</i>	Vergleich von eigenschafts- und simulationsorientierten Verfahrenskonzepten.....	23
<i>Tabelle 2:</i>	Übersicht über die 13 szenariospezifischen Leistungskennwerte (Verhaltensmasse) von ATLANTIS.....	63
<i>Tabelle 3:</i>	Beispiele für die Bewertung von Textinhalten.....	80
<i>Tabelle 4:</i>	Übersicht über die chronologisch geordneten Systemereignisse in ATLANTIS und die in dieser Arbeit verwendeten Abkürzungen	83
<i>Tabelle 5:</i>	Verteilung der Problemlösetypen in der Teilstichprobe 1	124
<i>Tabelle 6:</i>	Zusammenhang zwischen den szenariospezifischen Leistungskennwerten und Informationseröffnendem sowie Adäquatem Handeln.....	130
<i>Tabelle 7:</i>	Übereinstimmung zwischen neuem Auswertungsmodell und Expertenurteil für das Informationseröffnende Handeln in der Teilstichprobe 1	133
<i>Tabelle 8:</i>	Übereinstimmung zwischen neuem Auswertungsmodell und Expertenurteil für das Adäquate Handeln in der Teilstichprobe 1	134
<i>Tabelle 9:</i>	Übereinstimmung zwischen neuem Auswertungsmodell und Expertenurteil für das Informationseröffnende Handeln in der Teilstichprobe 2	136
<i>Tabelle 10:</i>	Übereinstimmung zwischen neuem Auswertungsmodell und Expertenurteil für das Adäquate Handeln in der Teilstichprobe 2	137
<i>Tabelle 11:</i>	Ergebnisse der Einzelvergleiche zur Prüfung von Verteilungsunterschieden zwischen der Teilstichprobe 1 bzw. 2 und der Gesamtstichprobe.....	182
<i>Tabelle 12:</i>	Verteilungskennwerte des Informationseröffnenden Handelns zu ausgewählten Spielzeitpunkten.....	189
<i>Tabelle 13:</i>	Verteilungskennwerte des Adäquaten Handelns zu ausgewählten Spielzeitpunkten	189
<i>Tabelle 14:</i>	Interkorrelationen zwischen den verschiedenen Spielzeitpunkten für das Informationseröffnende und Adäquate Handeln	190
<i>Tabelle 15:</i>	Medianwerte für die Leistungskennwerte in Abhängigkeit von den Prozessprofilen für das Informationseröffnende und Adäquate Handeln.....	191
<i>Tabelle 16:</i>	Übereinstimmung zwischen neuem Auswertungsmodell und Expertenurteil in Bezug auf die Problemlösetypen in der Teilstichprobe 1.....	192
<i>Tabelle 17:</i>	Übereinstimmung zwischen neuem Auswertungsmodell und Expertenurteil in Bezug auf die Problemlösetypen in der Teilstichprobe 2.....	192

Anhang B: Ergänzungen zum Szenario ATLANTIS

An dieser Stelle folgen ergänzende Angaben zum Szenario ATLANTIS. Abbildung 15 zeigt die Seekarte, die dem Probanden für Notizen zur Verfügung gestellt wird. Auf der nächsten Seite folgt eine Übersicht über alle Funktionen des Steuerungscomputers, vor dem der Proband bei der Bearbeitung von ATLANTIS sitzt. An dritter Stelle ist ein vollständiger Protokollbogen des szenariospezifischen Auswertungsansatzes wiedergegeben. Damit die einzelnen Bestandteile des Protokollbogens rasch gefunden werden können, sind deren Überschriften grau markiert.



Übersicht über die Funktionen des Steuerungscomputers Nautilus EA.4

Die grau markierten Überschriften 1-5 bezeichnen die verfügbaren Hauptmenüs (vgl. auch Abbildung 2, Seite 53).

1) Bedienungs-Hilfe

- Hilfe verlassen
- Steuerung Nautilus
- Video
- Magneto-Sonar
- Höhe/Tiefe: Oberfläche, Sinken, Steigen, Grund
- Bewegungsrichtung (8 Richtungen)
- Message lesen
- Protokoll/Message
 - Protokoll lesen
 - Protokoll erstellen
 - Message erstellen
- Nauti-Info
- System-Anzeigen: ?
 - System-Meldungen: ?
 - Energievorrat: ?
 - Monitor: ?
 - Koordinaten x/y: ?
 - Seekarte: ?
 - Meeres-, Tauchtiefe: ?

2) Steuerung Nautilus

- Video
- Magneto-Sonar
- Höhe/Tiefe: Oberfläche, Sinken, Steigen, Grund
- Bewegungsrichtung: 8 Richtungen

3) Message lesen

(Die letzte eingegangene Nachricht wird auf dem Monitor eingeblendet)

4) Protokoll/Message

- Protokoll(e) lesen
 - Vorhergehendes Protokoll
 - Nächstes Protokoll
- Protokoll erstellen
 - Protokoll speichern
 - Protokoll verwerfen
- Message erstellen
 - Senden an:
 - alle
 - Leitung
 - Tauchergruppe
 - Technik
 - Terminplaner
 - Service/Bordküche

5) Nauti-Info

- Protokolle
 - Strömung in Q/9, P/10
 - Neue Fluglinie???
 - Wrack statt Erz!
 - Sichtweite (Video)
 - Alka Selzer
- Geologie/Strömungen
 - Schlucht (Entstehung)
 - Schelfe
 - Strömung
 - Metalle (Vorkommen)
 - Salze/Sulfide/Phosphor
- Meeresbiologie
 - Polyphemus
 - Asselspinnen
 - Delphine (aktuell)
 - Chelonia mydas (akt)
 - Seegrasbetten
- Nautische Instrumente
 - Tauchgeräte (Tiefen)
 - Max. erreichte Tiefe
 - Unbemannte Fahrzeuge
 - Titanic (Entdeckung)
 - Rettungseinsatz
- Seemannsgarn
 - Allgemein
 - „Madagaskar“
 - „Oldsworth“ (aktuell)
 - Schiffsfahrtsunfall
 - Jack the Invisible
- Krankheit, Un-/Notfall
 - Krankheit
 - Unfälle (Verletzte)
 - Brand/Seenot
 - Sinkverhalten
 - Überlebenschancen

6) Programm beenden

(bzw. „Instruktion wiederholen“ im Training)

Protokollbogen des szenariospezifischen Auswertungsansatzes

Ausführen der Trainingsaufgabe, Dauer: 23 min 16 sec

```

-----
Lesen im Hilfesystem, Anteil Lesedauer: 26 %          *
Nautilus in mind. 3 relevanten Modi benutzt          *
Magneto-Sonareinsätze im Trainingsgebiet: 2          *
    Einsätze mit Meeresgrund: 2 (100%)              *
Video-Einsätze im Trainingsgebiet: 2                  *
    Einsätze ohne Meeresgrund: 0 ( 0%)               *
    Einsätze mit Meeresgrund: 2 (100%)              *
Metall im Scan-, nicht im Sichtbereich: 1 ( 50%)     *
Metall im Video-Sichtbereich: 0 ( 0%)

```

Dauer der Bearbeitung (ohne Trainingsaufgabe): 55 min 47 sec

```

-----
Lesen im Hilfesystem, Anteil Lesedauer: 0 %
Lesen im Info-System, Anteil Lesedauer: 2 %,
    relevante Informationen: 10 %
    neutrale Informationen: 89 %
    unterhaltende Informationen: 0 %

```

Vorgehen gebietsweise:.....

Tauchtiefe bei

langen Wegstrecken:.....
Scans:.....

Gebietswechsel bei Message

2.....
3.....
4.....

unnötige Wege:.....

```

Magneto-Sonareinsätze im gesamten Gebiet: 30          *
    Einsätze mit Meeresgrund: 27 ( 90%)              *
Video-Einsätze im gesamten Gebiet: 13                  *
    Einsätze ohne Meeresgrund: 2 ( 15%)               *
    Einsätze mit Meeresgrund: 11 ( 84%)              *
Metall im Scan-, nicht im Sichtbereich: 1 ( 7%)        *
Metall im Video-Sichtbereich: 10 ( 76%)              *

```

```

Magneto-Sonareinsätze im 50m-Bereich S/7: 24          *
    Einsätze mit Meeresgrund: 22 ( 91%)              *
Video-Einsätze im 50m-Bereich S/7: 11                  *
    Einsätze ohne Meeresgrund: 1 ( 9%)                *
    Einsätze mit Meeresgrund: 10 ( 90%)              *
Metall im Scan-, nicht im Sichtbereich: 1 ( 9%)        *
Metall im Video-Sichtbereich: 9 ( 81%)                *

```

```

Magneto-Sonareinsätze im 50m-Bereich R/9: 3           *
    Einsätze mit Meeresgrund: 3 (100%)                *
Video-Einsätze im 50m-Bereich R/9: 1                   *
    Einsätze ohne Meeresgrund: 0 ( 0%)                *
    Einsätze mit Meeresgrund: 1 (100%)                *
Metall im Scan-, nicht im Sichtbereich: 0 ( 0%)        *
Metall im Video-Sichtbereich: 1 (100%)                *

```

Kommunikation: Erstellen der Protokolle, Lesen und Erstellen der Messages

11 53 liest Trainings-Message "Erinnerung an Protokoll-Eintrag" (T 1) 112 sec. nach Eingang

15 12 liest Trainings-Message "Aufgabe beenden" (T 2) 3 sec. nach Eingang

16 21 >> Protokoll Nr. 1 406/305
S8
222
Keine Vorkommen durch Sonar festgestellt

16 24 liest Trainings-Message "Anweisung zum Auftanken" (T 3) 3 sec. nach Eingang

16 45 liest Trainings-Message "Anweisung zum Auftanken" (T 3) 24 sec. nach Eingang

17 21 => Message Nr. 1 wurde nicht versendet

19 29 liest Trainings-Message "Anweisung zum Auftanken" (T 3) 188 sec. nach Eingang

20 54 >> Protokoll Nr. 2 488/328
t9
100
unklar, Sonar vorkommen gesichtet

22 37 liest Trainings-Message "Anweisung zum Auftanken" (T 3) 376 sec. nach Eingang

23 13 => Message Nr. 2 Tanken
Technik
Bitte auftanken
-> an Technik

23 22 liest Trainings-Message "Anweisung zum Auftanken" (T 3) 421 sec. nach Eingang

23 26 liest Trainings-Message "Anweisung zum Auftanken" (T 3) 425 sec. nach Eingang

25 18 liest Absturzmeldung (Nr.1) 4 sec. nach Eingang

28 22 >> Protokoll Nr. 3 325/248
s8
0
Konzentration auf Tiefe bis 50m infolge
Überlebenschancen

31 56 >> Protokoll Nr. 4 157/248
s7
40
Sonar 3 Aufzeichnungen festgestellt
Weitere Abklärungen, annäherung nötig

32 38 ----- Gepäck auf Videobild, Koord.: 136/248 -----

33 3 liest Absturzmeldung (Nr.1) 469 sec. nach Eingang

33 50 >> Protokoll Nr. 5 136/248
s7
34
Video unklare Teile voraus festgestellt
Fahre näher

33 51 liest Absturzmeldung (Nr.1) 517 sec. nach Eingang

34 49 => Message Nr. 3 Wrackteile
in Abklärung
Taucher bereithalten für schnellen Einsa
tz
Sind weitere Erkenntnisse vorhanden?
-> an alle

34 56 liest Absturzmeldung (Nr.1) 582 sec. nach Eingang

35 48 liest Anfrage der Taucher (T) 3 sec. nach Eingang

36 49 => Message Nr. 4 Einsatz Taucher
Dringend
Einsatz bei 136/248, 2. Team Einsatzbere
itschaft
-> an Leitung

39 41 >> Protokoll Nr. 6 153/238
s7
35
Video nichts zu erkennen
Suche wird fortgesetzt

40 57 liest Message "Radar-Signale T/7-Q/9" (Nr.2) 6 sec. nach Eingang

41 56 liest Message "Radar-Signale T/7-Q/9" (Nr.2) 65 sec. nach Eingang

42 31 liest Message "Radar-Signale T/7-Q/9" (Nr.2) 100 sec. nach Eingang

42 33 liest Message "Radar-Signale T/7-Q/9" (Nr.2) 102 sec. nach Eingang

43 53 >> Protokoll Nr. 7 139/200
s7
20
Sonar Reagierte, Video keine Trümmer erk
ennbar
Suche an einem anderen Ort fortsetzten

43 54 liest Message "Radar-Signale T/7-Q/9" (Nr.2) 183 sec. nach Eingang

43 56 liest Message "Radar-Signale T/7-Q/9" (Nr.2) 185 sec. nach Eingang

44 26 => Message Nr. 5 Taucher
Frage
Sind die Taucher im Einsatz. Erkenntniss
e?
-> an Tauchergruppe

45 40 => Message Nr. 6 Leitung
Frage
Keine Funde in dieser Region. Wo soll di
e Suche fortgesetzt werden? Absprache Ei
nsatz Taucher!
-> an alle

46 32 >> Protokoll Nr. 8 124/215
s7
20
Keine Funde

48 48 >> Protokoll Nr. 9 103/231
s7
17
Sonar Reagierte. Keine Trümmer auf Video
erkennbar.
Suchkreis wird Richtung T7 ausgedehnt

48 54 liest Message "Frachtstücke gefunden in R/9" (Nr.3) 2 sec. nach
Eingang

49 31 liest Message "Frachtstücke gefunden in R/9" (Nr.3) 39 sec. nach
Eingang

50 30 => Message Nr. 7 Neuer Suchort
Dringend
Naut EA.4 Verlagert suche nach R9 gem. F
undort Kisten
-> an Leitung

52 12 => Message Nr. 8 Keine Funde
Geprüft
Sonareinsatz 165/169 Karte s7 keine Kont
akte
-> an Leitung

55 17 >> Protokoll Nr. 10 401/156
r8
52
Keine Sonarkontakt

55 56 liest Message "Radar-Signale = 2 Flugzeuge, verm. Flugzeug nördl.
T/7" (Nr.4) 3 sec. nach Eingang

56 12 liest Message "Radar-Signale = 2 Flugzeuge, verm. Flugzeug nördl.
T/7" (Nr.4) 19 sec. nach Eingang

56 48 liest Message "Radar-Signale = 2 Flugzeuge, verm. Flugzeug nördl.
T/7" (Nr.4) 55 sec. nach Eingang

59 2 >> Protokoll Nr. 11 441/145
r9
43
Sonar Kontakt. Video keine Trümmer.
Aufgrund neuer Message wird die Suche in
Karte S7 fortgesetzt

59 4 liest Message "Radar-Signale = 2 Flugzeuge, verm. Flugzeug nördl.
T/7" (Nr.4) 191 sec. nach Eingang

59 43 => Message Nr. 9 Neue Suche
Dringend
Aufgrund der neuen Message wird Naut EA.
4 die Suche im S7 neu aufnehmen.
-> an Leitung

63 9 ----- Flugzeug auf Videobild, Koord.: 165/190 -----

64 4 => Message Nr. 10 Fund Trümmer
Dringend
Trümmerteile bei 165/190 Tiefe 38m festg
estellt. Taucher Einsatz dringend.
-> an alle

64 10 liest Message "Flugzeug gefunden, Schnellboot unterwegs" (Nr.6) 2
sec. nach Eingang

64 24 liest Message "Flugzeug gefunden, Schnellboot unterwegs" (Nr.6) 16
sec. nach Eingang

65 46 >> Protokoll Nr. 12 159/196
s7
21
Mehrere Sonarkontakte lokalisiert. Video
Bestätigt Flugzeug
Gem. Message Rückkehr versorgungsschiff

68 7 liest Message "Flugzeug gefunden, Schnellboot unterwegs" (Nr.6) 239
sec. nach Eingang

69 47 => Message Nr. 11 Abklärung
Offen
Konnten die Taucher den Flugzeugfund bes
tätigen? Oder ist eine weitere Suche dur
ch Nau EA.4 erforderlich?
-> an Leitung

70 41 liest Message "Rettung vielversprechend, Dank an Nautilus" (Nr.7) 4
sec. nach Eingang

Kennwerte zur Trainingsaufgabe

Lesen im Hilfesystem:

3 18	H-Nautilus	offen für 68 sec.
4 26	H-zur Oberfläche	offen für 13 sec.
4 39	H-Grund-Automatik	offen für 12 sec.
4 51	H-Nautilus bewegen	offen für 31 sec.
5 22	H-Nautilus bewegen	offen für 9 sec.
5 31	H-Video-Kamera	offen für 18 sec.
5 49	H-Sonar	offen für 43 sec.
6 32	H-Message lesen	offen für 21 sec.
6 53	H-Protokoll/Message	offen für 49 sec.
10 8	H-Information	offen für 15 sec.
10 23	H-Information	offen für 1 sec.
10 24	H-Information	offen für 2 sec.
10 26	H-Message lesen	offen für 7 sec.
10 33	H-Message lesen	offen für 4 sec.
10 37	H-System-Anzeigen	offen für 2 sec.
10 39	H-System-Meldungen	offen für 10 sec.
10 49	H-Energie	offen für 7 sec.
10 56	H-Monitor	offen für 5 sec.
11 1	H-Koordinaten-Ausgabe	offen für 19 sec.
11 20	H-Seekarte	offen für 21 sec.
11 41	H-Meerestiefe	offen für 11 sec.

Gesamt-Lesedauer: 6 min 8 sec

Anteil Lesedauer total: 26 %

Anzeigen von Informationen:

Gesamt-Lesedauer: 0 min 0 sec

```
Anteil Lesedauer total:    0 %,
                           davon 0 % neutrale und
                           0 % unterhaltende Informationen
```

Protokolle und Messages:

```
16 21  >> Protokoll Nr. 1
17 21  => Message Nr. 1   wurde nicht versendet
20 54  >> Protokoll Nr. 2
23 13  => Message Nr. 2   -> an Technik
```

Einsatz Nautilus, Zeiten: erstmaliger Einsatz in diesem Bearbeitungsabschnitt

```
12 30 Nautilus bewegen
12 38 manuelles Sinken
13 31 automat. Abstand einschalten
20 57 automat. Steigen bis zur Oberfläche einschalten
```

	7	Magneto-Sonareinsätze (=100%), davon	
12	58	Einsätze ohne Meeresgrund:	1 (= 14 %)
13	53	Einsätze mit Meeresgrund:	6 (= 85 %)
	2	Video-Einsätze (=100%), davon	
		Einsätze ohne Meeresgrund:	0 (= 0 %)
19	16	Einsätze mit Meeresgrund:	2 (=100 %)
		Metall im Scan-, nicht im Sichtbereich:	1 (= 50 %)
		Metall im Video-Sichtbereich:	0 (= 0 %)

---> im Trainingsgebiet <---

	2	Magneto-Sonareinsätze im Trainingsgebiet (=100%), davon	
		Einsätze ohne Meeresgrund:	0 (= 0 %)
19	5	Einsätze mit Meeresgrund:	2 (=100 %)
	2	Video-Einsätze im Trainingsgebiet (=100%), davon	
		Einsätze ohne Meeresgrund:	0 (= 0 %)
19	16	Einsätze mit Meeresgrund:	2 (=100 %)
		Metall im Scan-, nicht im Sichtbereich:	1 (= 50 %)
		Metall im Video-Sichtbereich:	0 (= 0 %)

Kennwerte zur Flugzeugsuche

Lesen im Hilfesystem:

Gesamt-Lesedauer: 0 min 0 sec

Anteil Lesedauer total: 0 %

Anzeigen von Informationen:

0	26	Menü: Geologie/Strömung		
0	33	Menü: Meeresbiologie		
0	37	Menü: nautische Instrumente		
0	41	Max. erreichte Tiefe (1 Seite)	1 Seite(n)	3 sec
0	44	Menü: Seemannsgarn		
0	46	Menü: Protokolle		
0	48	Menü: Krankheit, Un-/Notfall		
0	53	Menü: Protokolle		
0	58	Wrack statt Erz (2 Seiten)	GANZ GELESEN! 2 Seite(n)	53 sec
1	51	Menü: Protokolle		
1	54	* Sichtweite (Video (2 Seiten)	1 Seite(n)	8 sec
33	54	Menü: Geologie/Strömung		
33	57	Strömung (2 Seiten)	2 Seite(n)	10 sec

Gesamt-Lesedauer: 1 min 14 sec

Anteil Lesedauer total: 2 %,
davon 10 % relevante,
89 % neutrale und
0 % unterhaltende Informationen

Protokolle und Messages:

2 2 liest Absturzmeldung (Nr.1) 4 sec. nach Eingang
5 6 >> Protokoll Nr. 3
8 40 >> Protokoll Nr. 4
9 47 liest Absturzmeldung (Nr.1) 469 sec. nach Eingang
10 34 >> Protokoll Nr. 5
10 35 liest Absturzmeldung (Nr.1) 517 sec. nach Eingang
11 33 => Message Nr. 3 -> an alle
11 40 liest Absturzmeldung (Nr.1) 582 sec. nach Eingang
12 32 liest Anfrage der Taucher (T) 3 sec. nach Eingang
13 33 => Message Nr. 4 -> an Leitung
16 25 >> Protokoll Nr. 6
17 41 liest Message "Radar-Signale T/7-Q/9" (Nr.2) 6 sec. nach Eingang
18 40 liest Message "Radar-Signale T/7-Q/9" (Nr.2) 65 sec. nach Eingang
19 15 liest Message "Radar-Signale T/7-Q/9" (Nr.2) 100 sec. nach Eingang
19 17 liest Message "Radar-Signale T/7-Q/9" (Nr.2) 102 sec. nach Eingang
20 37 >> Protokoll Nr. 7
20 38 liest Message "Radar-Signale T/7-Q/9" (Nr.2) 183 sec. nach Eingang
20 40 liest Message "Radar-Signale T/7-Q/9" (Nr.2) 185 sec. nach Eingang
21 10 => Message Nr. 5 -> an Tauchergruppe
22 24 => Message Nr. 6 -> an alle
23 16 >> Protokoll Nr. 8
25 32 >> Protokoll Nr. 9
25 38 liest Message "Frachtstücke gefunden in R/9" (Nr.3) 2 sec. nach Eingang
26 15 liest Message "Frachtstücke gefunden in R/9" (Nr.3) 39 sec. nach Eingang
27 14 => Message Nr. 7 -> an Leitung
28 56 => Message Nr. 8 -> an Leitung
32 1 >> Protokoll Nr. 10
32 40 liest Message "Radar-Signale = 2 Flugzeuge, verm. Flugzeug nördl. T/7" (Nr.4) 3 sec. nach Eingang
32 56 liest Message "Radar-Signale = 2 Flugzeuge, verm. Flugzeug nördl. T/7" (Nr.4) 19 sec. nach Eingang
33 32 liest Message "Radar-Signale = 2 Flugzeuge, verm. Flugzeug nördl. T/7" (Nr.4) 55 sec. nach Eingang
35 46 >> Protokoll Nr. 11
35 48 liest Message "Radar-Signale = 2 Flugzeuge, verm. Flugzeug nördl. T/7" (Nr.4) 191 sec. nach Eingang
36 27 => Message Nr. 9 -> an Leitung
40 48 => Message Nr. 10 -> an alle
40 54 liest Message "Flugzeug gefunden, Schnellboot unterwegs" (Nr.6) 2 sec. nach Eingang
41 8 liest Message "Flugzeug gefunden, Schnellboot unterwegs" (Nr.6) 16 sec. nach Eingang
42 30 >> Protokoll Nr. 12
44 51 liest Message "Flugzeug gefunden, Schnellboot unterwegs" (Nr.6) 239 sec. nach Eingang
46 31 => Message Nr. 11 -> an Leitung
47 25 liest Message "Rettung vielversprechend, Dank an Nautilus" (Nr.7) 4 sec. nach Eingang

Einsatz Nautilus, Zeiten: erstmaliger Einsatz in diesem Bearbeitungsabschnitt

5 21 Nautilus bewegen
6 31 manuelles Sinken
7 13 automat. Abstand einschalten
11 59 automat. Steigen bis zur Oberfläche einschalten
27 36 manuelles Steigen
30 32 drohende Kollision angezeigt

30 Magneto-Sonareinsätze (=100%), davon
 12 7 Einsätze ohne Meeresgrund: 3 (= 10 %)
 6 46 Einsätze mit Meeresgrund: 27 (= 90 %)
 13 Video-Einsätze (=100%), davon
 6 24 Einsätze ohne Meeresgrund: 2 (= 15 %)
 9 22 Einsätze mit Meeresgrund: 11 (= 84 %)
 Metall im Scan-, nicht im Sichtbereich: 1 (= 7 %)
 Metall im Video-Sichtbereich: 10 (= 76 %)

---> im 50m-Bereich S/7 <---

24 Magneto-Sonareinsätze im 50m-Bereich S/7 (=100%), davon
 12 7 Einsätze ohne Meeresgrund: 2 (= 8 %)
 6 46 Einsätze mit Meeresgrund: 22 (= 91 %)
 11 Video-Einsätze im 50m-Bereich S/7 (=100%), davon
 7 9 Einsätze ohne Meeresgrund: 1 (= 9 %)
 9 22 Einsätze mit Meeresgrund: 10 (= 90 %)
 Metall im Scan-, nicht im Sichtbereich: 1 (= 9 %)
 Metall im Video-Sichtbereich: 9 (= 81 %)

---> im 50m-Bereich R/9 <---

3 Magneto-Sonareinsätze im 50m-Bereich R/9 (=100%), davon
 Einsätze ohne Meeresgrund: 0 (= 0 %)
 31 7 Einsätze mit Meeresgrund: 3 (=100 %)
 1 Video-Einsätze im 50m-Bereich R/9 (=100%), davon
 Einsätze ohne Meeresgrund: 0 (= 0 %)
 34 46 Einsätze mit Meeresgrund: 1 (=100 %)
 Metall im Scan-, nicht im Sichtbereich: 0 (= 0 %)
 Metall im Video-Sichtbereich: 1 (=100 %)

ausgewählte Kennwerte

zur problembezogenen Aktivität und Effizienz

(Vorgaben in %)

36.8 % der potentiellen Fundorte in beiden Suchgebieten wurden gescannt

66.7 % der potentiellen Fundorte im linken Suchgebiet (S7) wurden gescannt

10.0 % der potentiellen Fundorte im rechten Suchgebiet (R9) wurden gescannt

68.1 % der Bearbeitungszeit für eingegangene Informationen wurden effizient genutzt

100.0 % der Bearbeitungszeit für abgerufene Informationen wurden effizient genutzt

69.7 % der Verschiebungen der Suchpunkte wurde effizient gestaltet

zur Problemerkennung und -lösung

89.2 % der Suchaktivitäten wurden in sinnvoller Tiefe (auf Grund) durchgeführt

86.7 % der Prüfkaktivitäten wurden in sinnvoller Tiefe (auf Grund) durchgeführt

83.3 % der Suchaktivitäten wurden in sinnvoller Tiefe und in sinnvollen Gebieten durchgeführt

83.0 % der Prüfkaktivitäten bezogen sich genau auf potentielle Funde in sinnvollen Gebieten

zum Kommunikationsfluss

95.7 % der Informationen wurden effizient aus-/weitergegeben

90.0 % der ankommenden Informationen wurden effizient aufgenommen/abgerufen

100.0 % der ankommenden Informationen wurden vollständig abgerufen

Anhang C: Ergänzungen zur Mikroverhaltensbeschreibung

Als erstes folgen genauere Angaben zur Datenbasis, die für die Mikroverhaltensbeschreibung zur Verfügung stand (Abbildung 16). Anschliessend wird letztere im Detail erläutert.

Spielzeit	X-Koordinate	Y-Koordinate	Sektor	Meerestiefe	Energievorrat	Tauchtiefe	Potentielle Sonar-Ziele	Abstand zum Meeresgrund	Zurückgelegte Distanz D (in Pixel)	D in Einheiten des Sonarradius	Richtung der Videokamera	Dauer in sec	Aktion bzw. Systemmeldung
00:25:03	32	24	S8	0	62	25	0 0 0 0 0	258	0	0	8	4	Nauti bereit
00:25:07	32	24	S8	0	62	25	0 0 0 0 0	258	0	0	8	3	10
00:25:10	32	24	S8	0	62	25	0 0 0 0 0	258	0	0	8	4	14
00:25:14	32	24	S8	0	62	25	0 0 0 0 0	258	0	0	8	4	K-M.m_not1:5
00:25:18	32	24	S8	0	62	25	0 0 0 0 0	258	0	0	8	36	Lese m_not1:1
00:25:54	32	24	S8	0	62	25	0 0 0 0 0	258	0	0	8	26	Lese m_not1:2
00:26:20	32	24	S8	0	61	25	0 0 0 0 0	258	0	0	8	33	Lese m_not1:3
00:26:53	32	24	S8	0	61	25	0 0 0 0 0	258	0	0	8	21	Lese m_not1:4
00:27:14	32	24	S8	0	61	25	0 0 0 0 0	258	0	0	8	16	Lese m_not1:5
00:27:30	32	24	S8	0	61	25	0 0 0 0 0	258	0	0	8	52	Proto.geöffnet
00:28:22	32	24	S8	0	61	25	0 0 0 0 0	258	0	0	8	5	Proto.Nr3 OK
00:28:27	32	24	S8	0	61	25	0 0 0 0 0	258	0	0	8	10	Nautilus
00:28:37	32	24	S8	0	61	25	0 0 0 0 0	258	0	0	8	62	Move!
00:29:39	18	24	S7	0	61	75	0 0 0 0 0	75	141	5.88	7	1	-60-
00:29:40	18	24	S7	0	61	75	0 0 0 0 0	75	0	0	7	7	Video
00:29:47	18	24	S7	0	61	75	0 0 0 0 0	75	0	0	7	11	Sinken
00:29:58	18	24	S7	4	61	75	0 0 0 0 0	27	0	0	7	4	Move!
00:30:02	18	24	S7	4	61	53	0 0 0 0 0	5	4	0.17	7	0	-2-
00:30:02	18	24	S7	4	61	53	0 0 0 0 0	5	0	0	7	0	Sonar
00:30:02	18	24	S7	4	61	53	0 0 0 0 0	5	0	0	7	23	Sonar:00000000
00:30:25	18	24	S7	4	61	53	0 0 0 0 0	5	0	0	7	4	Video
00:30:29	18	24	S7	4	61	53	0 0 0 0 0	5	0	0	7	1	Grund
00:30:30	18	24	S7	5	61	53	0 0 0 0 0	3	0	0	7	2	Grund
00:30:32	18	24	S7	5	61	53	0 0 0 0 0	3	0	0	7	16	Move!
00:30:48	15	24	S7	4	60	43	154 237	3	23	0.96	7	0	-14-
00:30:48	15	24	S7	4	60	43	154 237	3	0	0	7	0	Sonar
00:30:48	15	24	S7	4	60	43	154 237	3	0	0	7	13	Sonar:00000000
00:31:01	15	24	S7	4	60	43	0 0 0 0 0	3	0	0	7	55	Proto.geöffnet
00:31:56	15	24	S7	4	60	43	0 0 0 0 0	3	0	0	7	3	Proto.Nr4 OK
00:31:59	15	24	S7	4	60	43	0 0 0 0 0	3	0	0	7	1	Nautilus
00:32:00	15	24	S7	4	60	43	0 0 0 0 0	3	0	0	7	6	Move!
00:32:06	14	24	S7	4	60	44	154 237	3	10	0.42	7	0	-4-
00:32:06	14	24	S7	4	60	44	154 237	3	0	0	7	0	Sonar
00:32:06	14	24	S7	4	60	44	154 237	3	0	0	7	10	Sonar:00000000
00:32:16	14	24	S7	4	60	44	0 0 0 0 0	3	0	0	7	8	Move!
00:32:24	13	24	S7	3	60	37	154 237	3	11	0.46	7	0	-7-
00:32:24	13	24	S7	3	60	37	154 237	3	0	0	7	0	Sonar
00:32:24	13	24	S7	3	60	37	154 237	3	0	0	7	14	Sonar:10000011
00:32:38	13	24	S7	3	60	37	154 237	3	0	0	7	0	Netz
00:32:38	13	24	S7	3	60	37	154 237	3	0	0	7	25	Video
00:33:03	13	24	S7	3	60	37	0 0 0 0 0	3	0	0	7	5	Lese m_not1:1
00:33:08	13	24	S7	3	60	37	0 0 0 0 0	3	0	0	7	42	Proto.geöffnet
00:33:50	13	24	S7	3	60	37	0 0 0 0 0	3	0	0	7	1	Proto.Nr5 OK
00:33:51	13	24	S7	3	60	37	0 0 0 0 0	3	0	0	7	3	Lese m_not1:1
00:33:54	13	24	S7	3	60	37	0 0 0 0 0	3	0	0	7	55	Mess.geöffnet
00:34:49	13	24	S7	3	60	37	0 0 0 0 0	3	0	0	7	4	Mess.Nr3
00:34:53	13	24	S7	3	60	37	0 0 0 0 0	3	0	0	7	3	an alle
00:34:56	13	24	S7	3	60	37	0 0 0 0 0	3	0	0	7	14	Lese m_not1:1
00:35:10	13	24	S7	3	60	37	0 0 0 0 0	3	0	0	7	5	Nautilus
00:35:15	13	24	S7	3	60	37	0 0 0 0 0	3	0	0	7	6	Oberfl.
00:35:21	13	24	S7	0	59	37	0 0 0 0 0	37	0	0	7	2	Oberfl.
00:35:23	13	24	S7	0	59	37	0 0 0 0 0	37	0	0	7	0	Sonar
00:35:23	13	24	S7	0	59	37	0 0 0 0 0	37	0	0	7	10	Sonar:00000000
00:35:33	13	24	S7	0	59	37	0 0 0 0 0	37	0	0	7	3	Sinken
00:35:36	13	24	S7	1	59	37	0 0 0 0 0	25	0	0	7	0	Sonar
00:35:36	13	24	S7	1	59	37	0 0 0 0 0	25	0	0	7	9	Sonar:00000000
00:35:45	13	24	S7	1	59	37	0 0 0 0 0	25	0	0	7	3	K-M.m_tauch:1

Abbildung 16: Beispiel eines Verhaltensprotokolls. Dargestellt ist ein Ausschnitt aus dem Verhaltensprotokoll eines Probanden, der soeben das Training beendet hat.

Im Folgenden werden alle Interpretationsregeln der Mikroverhaltensbeschreibung aufgelistet. Die Reihenfolge der Darstellung orientiert sich an den im Text erwähnten Handlungskategorien Lesen, Schreiben und Suchen/Prüfen (vgl. Kapitel 3.2.1.1) sowie dem Aufbau des Steuerungscomputers Nautilus EA.4 (vgl. Anhang B). Für jede Handlungskategorie erfolgt zuerst eine Übersicht über alle möglichen Textausgaben sowie den für die Bewertung notwendigen Erklärungen dazu. Die anschließende Auflistung der jeweiligen Bewertungskriterien ist (sofern inhaltlich plausibel) reduziert auf die Bedingungen für die Ausprägungen 1 und 0.5 (= viel bzw. wenig Infozuwachs) sowie +1 und -1 (= sinnvoll bzw. nicht sinnvoll). Aktionen, die diese Bedingungen nicht erfüllen, werden mit Null bewertet (kein Informationszuwachs bzw. wenig sinnvoll).

Die Codierung beginnt mit dem Lesen der Übungs-Instruktion und endet mit der letzten Aktion vor der Systemmeldung „Ende“.

KATEGORIE LESEN

Übersicht über die Textausgaben mit Erklärungen

Für jede Textausgabe ist die **minimale Lesedauer (= min. LD)** angegeben, das heisst die Zeitdauer, während der eine Seite geöffnet sein muss, damit sie als gelesen gilt (Angabe in Sekunden, sofern nicht anders vermerkt).

ÜBUNGSINSTRUKTION

Übungs-Inst. - Ende-Inst.	2 Minuten
Übungs-Inst-W	10 (Wiederholung der Übungsinstruktion)

BEDIENUNGSHILFE

Help On	10 (da Menü-Übersicht und erklärender Text)
H-Nautilus	10 (da Menü-Übersicht und erklärender Text)
H-Video, H-Sonar	je 10
H-Move	10
H-Sinken, H-Steigen	je 5
H-Grund, H-Oberfläche	je 5
H-Mess. lesen	10
H-Prot./Mess.	10
H-Prot. lesen	10
H-Prot. erstellen	20 (da 2 Seiten, 2 Klicks)
H-Mess. erstellen	20 (da 2 Seiten, 2 Klicks)
H-Information	10
H-System-Anz.	2 (da nur Menü-Übersicht, kein Text)
H-Syst.-Meld., H-Energie	je 5
H-Monitor, H-Koord.	je 5
H-Seekarte, H-Tiefen	je 5
Help Off	keine Bewertung, da kein Text

EINGEHENDE MELDUNGEN

An erster Stelle ist die Systemmeldung aufgeführt, die der Proband erhalten kann, an zweiter Stelle die dazugehörige Aktion (Lesen der Meldung).

Ü-M.m_t1:1, lese m_t1:1	10 (Trainingsmeldung T1: Protokollerinnerung)
Ü-M.m_t2:1, lese m_t2:1	10 (Trainingsmeldung T2: Protokollaufforderung)
Ü-M.m_t3:2, lese m_t3:1(-2)	je 10 (Trainingsmeldung T3: Auftanken!)
Ü-M.m_t4:1, lese m_t4:1	10 (Trainingsmeldung T4: Positionskorrektur)
K-M.m_not1:5, lese m_not1:1(-5)	je 10 für alle 5 Seiten (Meldung M1: Notfall)
K-M.m_tauch:1, lese m_tauch:1	10 (Meldung T: Anfrage der Taucher)
K-M.m_tauchA:1, lese m_tauchA:1	5 (Meldung A: Antwort der Taucher)

K-M.m_not2:2, lese m_not2:1(-2)	10, für 2. Seite 5 (Meldung M2: Radarsignale)
K-M.m_fracht:1, lese m_fracht:1	10 (Meldung M3: Frachtstücke)
K-M.m_7not3:2, lese m_7not3:1(-2) (bzw. K-M.m_9not3:2)	10, für 2. Seite 5 (Meldung M4: Ort sensitiv)
K-M.m_end1:2, lese m_end1:1(-2)	10, für 2. Seite 5 (Meldung M5, Ende 1: Fischfänger)
K-M.m_end2:1, lese m_end2:1	10 (Meldung M6, Ende 2-1: Flugzeug lokalisiert)
K-M.m_end3:2, lese m_end3:1(-2)	10, für 2. Seite 5 (Meldung M7, Ende 2-2: Dank)
Keine Mess. vorhanden	keine min. LD (Systemmeldung, wenn Proband auf Funktion „Message lesen“ klickt und bisher, d.h. innerhalb der ersten 10 Minuten, noch keine Nachricht verschickt wurde)

DATENBANK NAUTI-INFO

Alle Textausgaben sind Zahlen. Zur Information wird in Klammer die Funktionsbeschriftung (das, was der Proband am Computer sieht) angegeben. Die einzelnen Seiten sind ein- oder mehrseitig, wobei hier zur Vereinfachung nur die Textausgabe der ersten Seite aufgeführt wird.

Im Menü 10 können nur die Textausgaben „Nächste Seite“ bewertet werden, die Textausgaben 11 bis 15 enthalten keine Informationen und werden nicht bewertet.

Relevante Seiten: 12, 14, 45, 64, 65; unterhaltende Seiten: 52, 53, 54, 55

0 (= Nauti-Info anklicken)	5 (da Menü-Übersicht, kein Text)
10 (= Protokolle)	10 (da Text und Menü-Übersicht)
11 (= Strömung in Q/9, P/10)	5 für „Nächste Seite“
12 (= neue Fluglinie???)	10 für „Nächste Seite“
13 (= Wrack statt Erz!)	5 für „Nächste Seite“
14 (= Sichtweite, Video)	10 für „Nächste Seite“
15 (= Alka Selzer)	10 für „Nächste Seite“
20 (= Geologie/Strömungen)	10 (da Text und Menü-Übersicht)
21 (= Schlucht, Entstehung)	5
22 (= Schelfe)	je 10 (mehrseitig)
23 (= Strömung)	je 10 (mehrseitig)
24 (= Metalle, Vorkommen)	10
25 (= Salze/Sulfide/Phosphor)	je 10 (mehrseitig)
30 (= Meeresbiologie)	10 (da Text und Menü-Übersicht)
31 (= Polyphemus)	je 10 (mehrseitig)
32 (= Asselspinnen)	je 10 (mehrseitig)
33 (= Delphine)	2
34 (= Chelonia mydas)	5
35 (= Seegrasbetten)	5
40 (= Naut. Instr.)	10 (da Text und Menü-Übersicht)
41 (= Tauchgeräte, Tiefen)	je 10 (mehrseitig)
42 (= Max. erreichte Tiefe)	10
43 (= Unbemannte Fahrzeuge)	10
44 (= Titanic, Entdeckung)	5
45 (= Rettungseinsatz)	je 10 (mehrseitig)
50 (= Seemannsgarn)	10 (da Text und Menü-Übersicht)
51 (= Allgemein)	10
52 (= „Madagaskar“)	10
53 (= „Oldsworth“)	je 10 (mehrseitig)
54 (= Schiffsfahrtsunfall)	je 10 (mehrseitig)
55 (= Jack the Invisible)	10
60 (= Krankheit, Un-/Notfall)	10 (da Text und Menü-Übersicht)
61 (= Krankheit)	10, für 2. Seite 5
62 (= Unfälle, Verletzte)	5
63 (= Brand/Seenot)	10, für 2. Seite 5
64 (= Sinkverhalten)	je 10 (mehrseitig)
65 (= Überlebenschancen)	10

Allgemeine Bewertungskriterien

- Als 1. Lesen wird das 1. Anklicken einer Seite bezeichnet, sofern min. LD erreicht wurde*.
- Für das Zählen von **Wiederholungen (= W)** gilt: fortlaufend jedes Anklicken durchnummerieren, sofern Seite mindestens 3 Sekunden lang geöffnet ist.
- Bei Doppel- und Mehrfachklicks (= unmittelbar wiederholtes Aufrufen der gleichen Seite) werden die Zeiten der einzelnen Seiten zusammengezählt und in Summe mit der min. LD verglichen.
- Als **Übersichtsseiten** zählen Seiten, die einen Überblick über das verfügbare Menü geben. In der Datenbank Nauti-Info sind dies die Textausgaben 0, 10, 20, 30, 40, 50, 60, in der Bedienungshilfe die Textausgaben Help-On, H-Nautilus, H-Prot./Mess., H-System-Anz.

Erstmaliges Lesen*	Info: +1 wenn LD \geq min. LD
	Ad: +1 wenn LD \geq min. LD
Wiederholtes Lesen*	Info: +0.5 wenn 1. Wiederholung und LD \geq min. LD
	Ad: -1 wenn > 1 Wiederholung und LD \geq min. LD

*Ausnahme: Wird beim 1. Anklicken die min. LD nicht erreicht (Info/Ad = 0/0), so zählt das 2. Anklicken (die 1. Wiederholung) als +1/+1, jede weitere Wiederholung zählt entsprechend den Kriterien 0/-1 (sofern LD \geq min. LD).

Zusatzbedingungen

DATENBANK NAUTI-INFO

Relevante Seiten	analog allg. Kriterien
Neutrale Seiten	Info: analog allg. Kriterien
	Ad: 0 unabh. von LD; W analog allg. Kriterien
Unterhaltende Seiten	Info: 0 unabh. von LD und W
	Ad: -1 wenn LD \geq min. LD (auch beim 1. Mal)
Übersichtsseiten	Info: +1 wenn LD \geq min. LD
	+0.5 wenn LD < min. LD
	Ad: +1 beim 1. Anklicken (unabhängig von LD)
	W jeweils analog allg. Kriterien, aber ohne Ausnahme

EINGEHENDE MELDUNGEN

Verzögerung beim Lesen: wenn (in Testphase) grössere Zeitdifferenz zwischen Ankunft der Meldung und Lesezeitpunkt; dazwischen sind Aktionen, die der Proband auch unterbrechen könnte (z.B. mehrere Sonareinsätze).

Nichtlesen von Meldungen	Ad: -1 wenn während Testphase
Verzögerungen	Ad: -1 (für die 1. Seite, danach analog allg. Krit.)
„lese m_t2:1 bzw. m_t4:1“	Info: 0 unabh. von LD
	Ad: -1 beim 1. Anklicken (unabhängig von LD)
	W jeweils analog allg. Kriterien, aber ohne Ausnahme
„keine Mess. vorhanden“	Info: +1 beim 1. Anklicken
	Ad: +1 beim 1. Anklicken

ÜBUNGSINSTRUKTION

„Übungs-Inst.“ - „Ende Inst.“	Info: +1 wenn LD \geq min. LD
	Ad: +1 wenn LD zwischen 2 und 5 Minuten
	0 wenn LD < 2 Minuten oder LD > 5 Minuten
„Übungs-Inst-W“	Info: +0.5 beim 1. Anklicken und wenn LD \geq min. LD
	Ad: +1 beim 1. Anklicken und wenn LD \geq min. LD
	-1 wenn LD > 3 Minuten

KATEGORIE SCHREIBEN

Übersicht über die Textausgaben mit Erklärungen

Proto. geöffnet	Modus Protokoll eröffnet
Proto. Nr. x OK	Protokoll gespeichert
Proto. Nr. x NOK	Protokoll verworfen
Mess. geöffnet	Modus Meldung eröffnet
Mess. Nr. x, an	Text verfasst und Meldung verschickt an.... (fehlt die Textausgabe „an...“ wurde Meldung nicht verschickt)
Protok. verlangt	Systemmeldung (Aufforderung zum Protokolleintrag)

Formale Bewertungskriterien in Bezug auf Ad (ohne Berücksichtigung des Inhaltes)

„Proto. Nr. x NOK“	-1
„Mess. Nr. x“, ohne „an...“	-1
„Proto. Nr. x OK“	+1 wenn aufgrund von „Protok. verlangt“
Ineffizientes Schreiben	-1 wenn > 2 Minuten pro Texteinheit; codieren bei „Proto./Mess. geöffnet“, die inhaltliche Bewertung erfolgt separat unter „Proto./Mess. Nr.x“

Inhaltliche Bewertungskriterien in Bezug auf Ad

Für die Beurteilung der formalen Korrektheit, der inhaltlichen Verständlichkeit und der Bedeutung im aktuellen Handlungskontext, vgl. Kapitel 3.2.1.4

„Proto. Nr. x OK“ und „Mess. Nr. x, an...“	+1 wenn formal korrekt (Modus, korrekte Angaben) und verständlich und im aktuellen Kontext sinnvoll, d.h. wenn inhaltliche Kriterien erfüllt sind -1 wenn im aktuellen Kontext nicht sinnvoll oder wenn formal nicht korrekt (falscher Modus, falsche Angaben) oder wenn nicht verständlich
Zusatzbedingung für „Proto. Nr. 1 OK“	+1 für kreative, besondere Idee (Zusatzpunkt) +1 wenn Nautilus in T/9 ist bzw. war und inhaltliche Kriterien erfüllt sind +1 wenn aufgrund von Trainingsmeldung T2 (Protokoll-aufforderung) und inhaltliche Kriterien erfüllt sind -1 wenn Nautilus nicht in T/9 ist bzw. war oder wenn inhaltliche Kriterien nicht erfüllt sind

KATEGORIE SUCHEN/PRÜFEN

Übersicht über die Textausgaben mit Erklärungen

Nautilus	Menü Steuerung wird aufgerufen
Crash!	Systemmeldung (weist auf bevorstehende Kollision hin)
Move! -xx-	Aktivierung des Modus „Fortbewegung in horizontaler Richtung“ während der angegebenen Zeitdauer „xx“ (in Sekunden); die zurückgelegte Distanz wird in Pixel oder in Einheiten des Sonarradius (= r) ausgegeben (vgl. Verhaltensprotokoll)
Sinken, Steigen	Aktivierung der Modi Sinken bzw. Steigen, oder Ausschalten der Steuerautomatik Grund bzw. Oberfl.
Grund, Oberfl.	Ein- bzw. Ausschalten der Steuerautomatik
Sonar (= S)	Sonaraufnahme wurde angefordert
Sonar: 00000000	Resultat der Sonaraufnahme
Video (= V)	Videobild wurde angefordert

Netz bzw. Wrack

Systemmeldungen; Proband sieht auf angefordertem Videobild ein Gepäckstück (Netz) bzw. einen Teil des havarierten Flugzeuges (Wrack)

Korrekte Interpretation der Textausgaben Grund, Oberfl. und Sonar

„Grund - Grund“

= Einschalten der Grund-Automatik, wenn beim 2. Grund ≤ 3 Meter über Grund; automatische Abstandseinhaltung wird ausgeschaltet durch Sinken, Steigen oder Oberfl.= Doppelklick, wenn > 3 Meter über Grund, oder wenn automatische Abstandseinhaltung bereits eingeschaltet ist (Klick hat keine Wirkung)

„Grund“

= Einschalten der Grund-Automatik, wenn bereits auf Grund (≤ 3 Meter über Grund)

= Keine Wirkung, wenn Automatik bereits eingeschaltet ist und erneut Grund gedrückt wird

= Ausschalten der Oberfl.-Automatik, wenn davor Oberfl.-Automatik eingeschaltet wurde

„Oberfl. - Oberfl.“

= Durch Automatik an Meeresoberfläche angekommen, wenn Tauchtiefe beim 2. Oberfl. = 0

= Doppelklick, wenn Tauchtiefe > 0 , oder wenn bereits an Oberfläche (Tauchtiefe = 0; Klick hat keine Wirkung)

„Oberfl.“

= Einschalten der Oberfl.-Automatik (wird ausgeschaltet durch Sinken, Steigen, Grund)

= Keine Wirkung, wenn bereits an Oberfläche (Tauchtiefe = 0)

= Ausschalten der Grund-Automatik, wenn davor Grund-Automatik eingeschaltet wurde

„Sonar: 0000000“

= S ohne Meeresgrund, wenn > 20 Meter über Grund (= S_{og})= S mit Meeresgrund, aber ohne Metallfund, wenn ≤ 20 Meter über Grund (= S_{mg_o})= S mit Meeresgrund, mit Metallfund im Scan-, aber nicht im Videosichtbereich, wenn ≤ 20 Meter über Grund und unter „Sonar-Ziele“ Koordinaten angegeben sind (= S_{mg_ms})

„Sonar: 01001000“

= S mit Meeresgrund, mit Metallfund im Videosichtbereich (= S_{mg_mv}; eine Eins steht für Metallvorkommen im entsprechenden Sektor des Sonarradius)*Bewertungskriterien für die Fortbewegung von Nautilus in Bezug auf Ad:*

„Crash!“

-1

„Move!/Sinken/Steigen“

werden als reine Fortbewegung nicht bewertet*, aber berücksichtigt für die Bewertung nachfolgender S/V

*Ausnahme für „Move!“:

+1 wenn „Move!“ $> 3r$ Achtung: wenn „Move!($>3r$)-andere-Aktionen-Move!($>3r$)“, dann wird nur 1. Move bewertet; bei „Move!-Sinken/Steigen-Move!“ Distanzen addieren und falls zusammen $> 3r$, dann wird 2. Move mit +1 bewertet

„Grund (- Grund)“

+1 wenn damit Automatik eingeschaltet wird und nachher S/V folgen

-1 wenn Automatik sofort wieder ausgeschaltet wird, z.B. „Grund(-Grund)-Sinken/Steigen/Oberfl.“

„Sinken/Steigen/Oberfl.“

wenn damit Grund-Automatik ausgeschaltet wird:

(alleine stehend, ohne Grund unmittelbar davor)

-1 wenn nachher weitere S/V folgen

+1 wenn unmittelbar danach „Move!“ $> 1r$ (Automatik wird ausgeschaltet, um zu fahren)

„Oberfl. (- Oberfl.)“

-1 wenn bereits an Oberfläche und erneut „Oberfl.“ gedrückt wird (für Mehrfachklicks nur 1 Mal zählen)

-1 wenn an Oberfläche angekommen und unmittelbar danach „sinken/Grund“ ODER S/V

+1 wenn an Oberfläche angekommen und unmittelbar danach „Move!“ $> 3r$ ODER kurz vor dem Auftanken (sofern kein S/V dazwischen)

Bewertungskriterien für den Einsatz des Sonargerätes:

„Sonar: 00000000“	Info:	+1 wenn ≤ 20 Meter über Grund (S _{mg})
	Ad:	+1 wenn ≤ 20 Meter über Grund (S _{mg})
		-1 wenn > 20 Meter über Grund (S _{og})
„Sonar: 01100010“	Info/Ad:	+1/+1

Zusatzbedingungen für alle S_{mg}

- a) *Sonar-Folgen*: mehrere Sonareinsätze hintereinander, ohne Aktionen dazwischen, die nicht die Steuerung betreffen, z.B. „S-Move!-S“ oder „S-V-Sinken/Steigen-Move!-S“; *Beurteilung der Effizienz der Verschiebung x*

nach S _{mg_o}	Info:	+1 wenn $x \geq 0.8r$
	Ad:	+1 wenn $1.2r \leq x \leq 2r$ (optimale Verschiebung)
		-1 wenn $x < 0.8r$
nach S _{mg_ms/mv}	Info:	+1 unabh. von Verschiebung
	Ad:	+1 wenn $x \leq 1r$ (optimale Verschiebung)
		-1 wenn $x > 1r$

- b) *Sonar-Reihen*: mehrere Sonareinsätze hintereinander, mit anderen Aktionen (lesen, schreiben) dazwischen; *Beurteilung von Wiederholungen*

keine Änderung*	Info/Ad:	0/-1
nur Änderung der Tiefe	Info/Ad:	0/-1
nur Änderung der Richtung	Info/Ad:	+1/+1 wenn nach S _{mg_mv}
		0/-1 wenn nach S _{mg_o/ms}
		+1/+1 wenn nach S _{mg_ms} und nachfolgender S = S _{mg_mv}
> 5 S mit „Move!“ < 5 Pixel	Info/Ad:	0/-1

Bewertungskriterien für den Einsatz der Videokamera:

Eine Videoaufnahme gilt als **korrekt**, wenn die drei folgenden Bedingungen erfüllt sind:

- Distanz zum Meeresgrund < 4 Meter
- Vor Videoeinsatz erfolgte S_{mg_mv} (Sonar: 01001000), und zwischen S und V erfolgte keine Ortsveränderung* (Änderungen der Richtung oder der Tauchtiefe sind erlaubt, sofern a und c erfüllt sind)
- Metallfund liegt im Videosichtbereich, d.h. Richtung von Nautilus stimmt mit Treffer-Sektoren überein (für Sonar-Beispiel in b müsste die Richtung 2 oder 5 sein)

„Video“	Info/Ad:	+1/+1 wenn alle Bedingungen erfüllt
		0/-1 wenn 1 Bedingung nicht erfüllt

Zusatzbedingungen und Ausnahmen

„Netz/Wrack - Video - Video“	Info:	+1 für beide Video (auch wenn V \neq korrekt)
	Ad:	-1 für 1. Video, wenn dieser nicht korrekt ist
		+1 für 1. Video, wenn dieser korrekt ist
		+1 für 2. Video (auch wenn V \neq korrekt)
Bedingung b) ist verletzt	Info/Ad:	+1/0 wenn S _{mg_mv/ms} -Move!-V-S _{mg_mv} und letzterer V bestätigt**
keine Änderung*	Info/Ad:	0/-1
> 3 V mit nur Änd. Richtung	Info/Ad:	0/-1

* keine Änderung bedeutet: gl. Ort (bzw. „Move!“ < 2 Pixel) und gl. Richtung und gl. Tiefe

** zwischen V und bestätigendem S_{mg_mv} dürfen andere Aktionen (aber kein „Move!“) vorkommen

Anhang D: Ergänzungen zu den Ergebnissen

Ergänzungen zu Kapitel 4.1	181
Ergänzungen zu Kapitel 4.2.1	184
Ergänzungen zu Kapitel 4.2.2	189
Ergänzungen zu Kapitel 4.4.1	191
Ergänzungen zu Kapitel 4.4.2 und Kapitel 4.4.3	192

Ergänzungen zu Kapitel 4.1

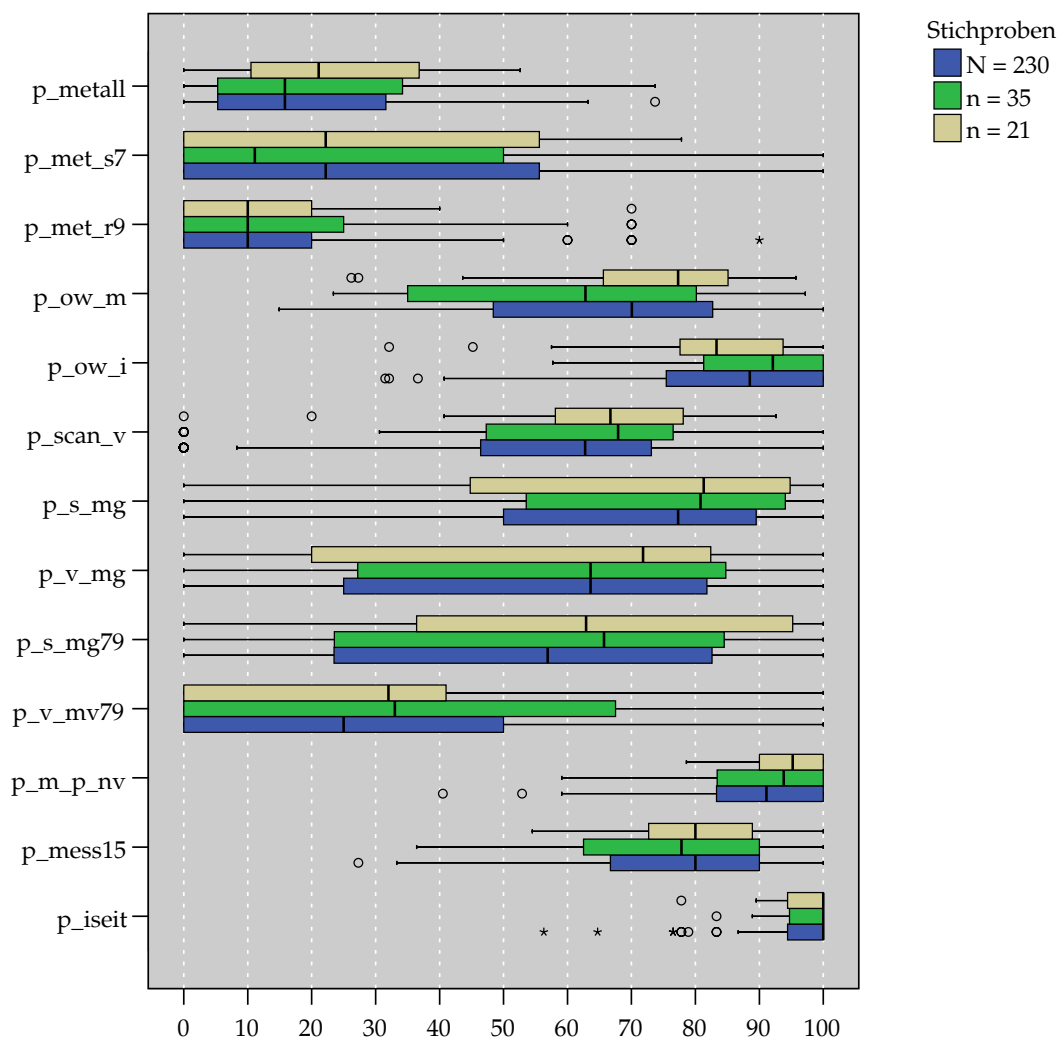


Abbildung 17: Verteilungen der 13 Leistungskennwerte in den untersuchten Stichproben. Die Verteilungen für die Gesamtstichprobe ($N = 230$) sind blau, die Verteilungen für die Teilstichprobe 1 ($n_1 = 35$) sind grün, und die Verteilungen für die Teilstichprobe 2 ($n_2 = 21$) sind beige dargestellt. Es kommen keine fehlenden Werte vor, ausser im Kennwert p_ow_i ($N = 196$, $n_1 = 28$, $n_2 = 17$).

Tabelle 11: Ergebnisse der Einzelvergleiche zur Prüfung von Verteilungsunterschieden zwischen der Teilstichprobe 1 ($n_1 = 35$) bzw. 2 ($n_2 = 21$) und der Gesamtstichprobe ($N = 230$).

Merkmale der Problembearbeitung	Teilstichprobe 1		Teilstichprobe 2	
	KS-Z*	p**	KS-Z*	p**
Leistungskennwerte				
p_metall	.49	.97	.51	.96
p_met_s7	.53	.94	.36	1.00
p_met_r9	.46	.98	.68	.75
p_ow_m	.79	.57	1.07	.20**
p_ow_i (N = 196, $n_1 = 28$, $n_2 = 17$)	.81	.53	.78	.59
p_scan_v	.74	.65	.81	.54
p_s_mg	.77	.59	.93	.36
p_v_mg	.45	.99	.73	.66
p_s_mg79	.51	.96	.97	.31
p_v_mv_79	.98	.29	.81	.53
p_m_p_nv	.61	.86	.94	.34
p_mess15	.40	1.00	.57	.90
p_iseit	.33	1.00	.28	1.00
Spielzeit				
- im Training	.46	.99	.62	.84
- in Testphase	.77	.59	.58	.89
Anzahl Protokolle				
- im Training	.50	.96	.52	.95
- in Testphase	1.00	.27	.54	.93
Anzahl Meldungen				
- im Training	.31	1.00	.54	.93
- in Testphase	.39	1.00	.26	1.00
Anzahl Sonareinsätze				
- im Training	.54	.93	.48	.93
- ohne Grund	.29	1.00	.44	.99
- mit Grund	.49	.97	.37	1.00
- in Testphase	.94	.34	.50	.96
- ohne Grund	.92	.36	.44	.99
- mit Grund	.73	.67	.30	1.00
Anzahl Videoeinsätze				
- im Training	.57	.91	.85	.47
- ohne Meeresgrund	.47	.98	.81	.53
- mit Meeresgrund	.44	.99	.54	.93
- mit Metall im Scanbereich	.15	1.00	.49	.97
- mit Metall im Videosichtbereich	.20	1.00	.28	1.00
- in Testphase	1.13	.15**	1.15	.14**
- ohne Meeresgrund	.98	.29	.33	1.00
- mit Meeresgrund	1.09	.18**	1.14	.15**
- mit Metall im Scanbereich	1.42	.04**	.94	.34
- mit Metall im Videosichtbereich	.50	.96	.81	.53

* Prüfgrösse Z des Kolmogorov-Smirnov-Test.

** Um den β -Fehler möglichst klein zu halten, wurde ein Signifikanzniveau von 25% gewählt. Signifikante Ergebnisse sind fett gedruckt und mit ** versehen.

Tabelle 11 (Fortsetzung): Ergebnisse der Einzelvergleiche zur Prüfung von Verteilungsunterschieden zwischen der Teilstichprobe 1 ($n_1 = 35$) bzw. 2 ($n_2 = 21$) und der Gesamtstichprobe ($N = 230$).

Merkmale der Problembearbeitung	Teilstichprobe 1		Teilstichprobe 2	
	KS-Z*	p**	KS-Z*	p**
Lesezeit Hilfesystem				
- im Training	1.21	.11**	.58	.89
- in Testphase	.51	.96	.68	.74
Lesezeit Hilfesystem, prozentualer Anteil				
- im Training	1.24	.09**	.53	.94
- in Testphase	.22	1.00	.51	.96
Lesezeit Datenbank Nauti-Info				
- im Training	.50	.96	.66	.78
- neutrale Informationen	.32	1.00	.45	.99
- unterhaltende Informationen	.22	1.00	.11	1.00
- in Testphase	.76	.61	.54	.94
- neutrale Informationen	.59	.88	.40	1.00
- unterhaltende Informationen	.10	1.00	.36	1.00
- relevante Informationen	.90	.40	.64	.81
Lesezeit Datenbank Nauti-Info, prozentualer Anteil				
- im Training	.55	.92	.68	.75
- in Testphase	.43	.99	.42	.99

* Prüfgrösse Z des Kolmogorov-Smirnov-Test.

** Um den β -Fehler möglichst klein zu halten, wurde ein Signifikanzniveau von 25% gewählt. Signifikante Ergebnisse sind fett gedruckt und mit ** versehen.

Ergänzungen zu Kapitel 4.2.1

Abbildung 18 zeigt die zeitliche Entwicklung des Informationseröffnenden und Adäquaten Handelns für alle Probanden der Teilstichprobe 1. Aus Platzgründen sind die Graphiken auf ein Minimum reduziert, so dass wohl die Zeitreihen, nicht aber die dazugehörigen Systemereignisse und vom Probanden erstellten Texteinheiten (gelbe Marken) gut erkenntlich sind (für Erläuterungen vgl. Einleitung zu Kapitel 4.2).

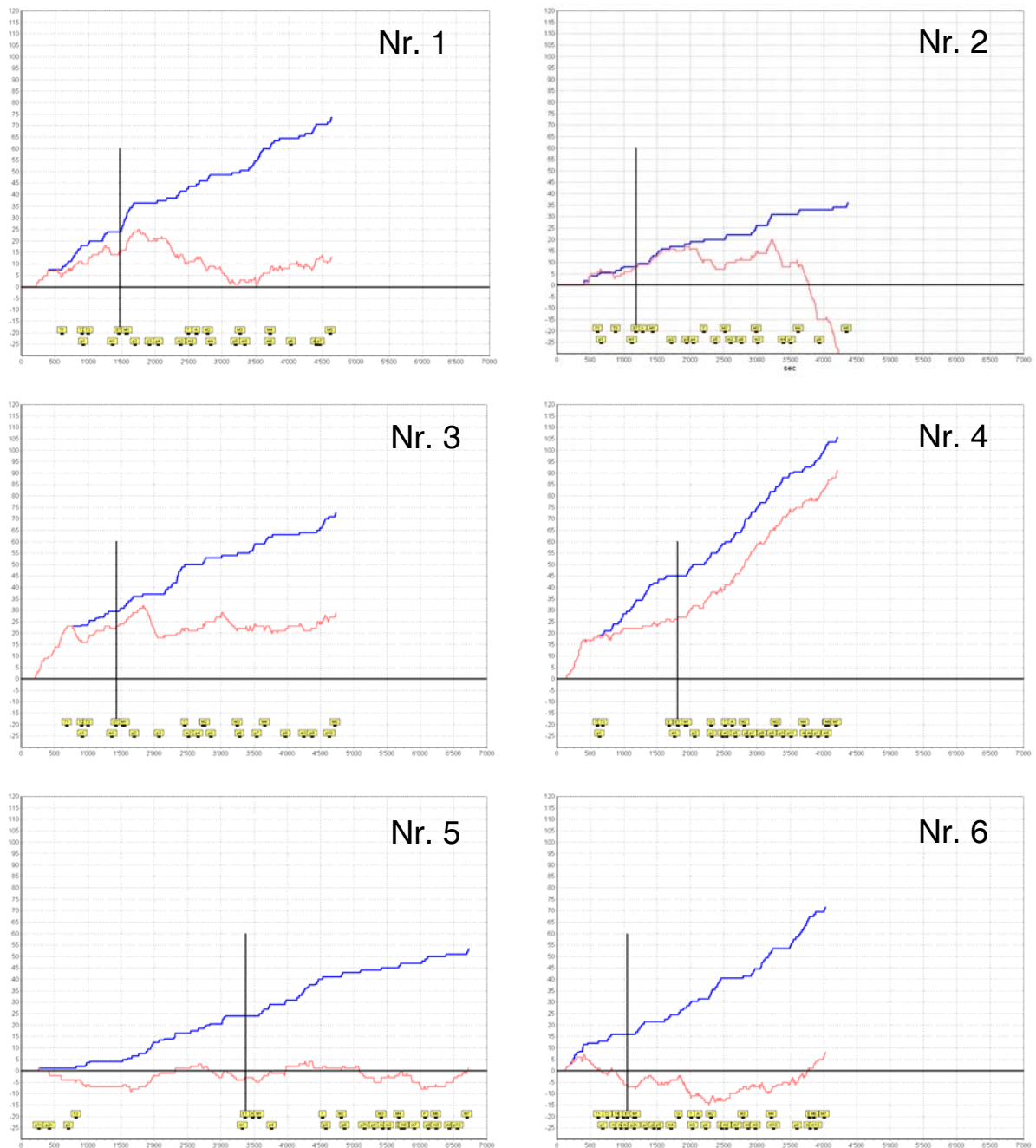


Abbildung 18: Darstellung der Zeitreihen für die Teilstichprobe 1.

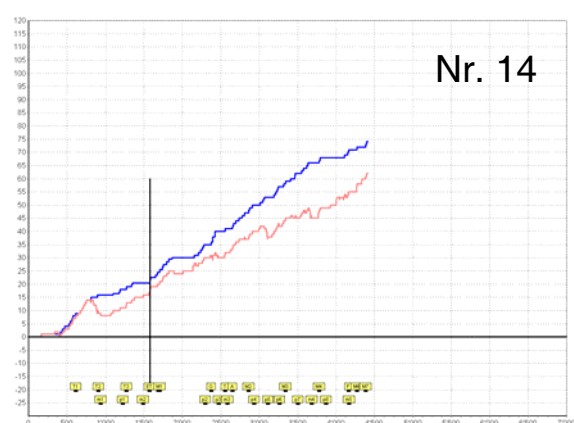
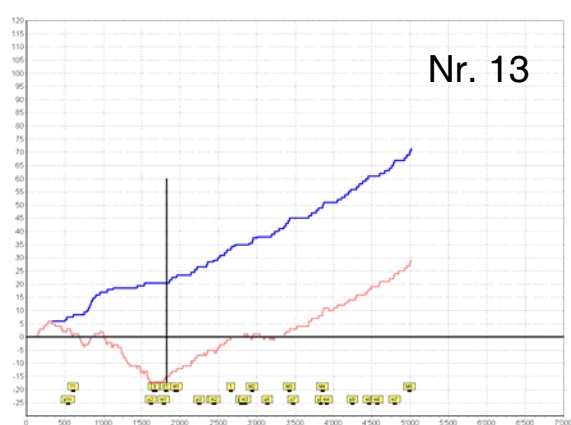
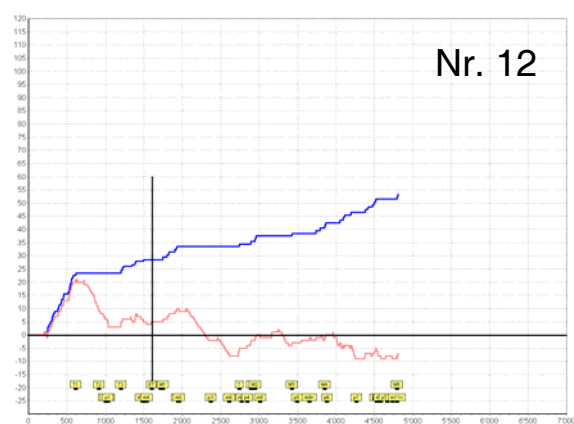
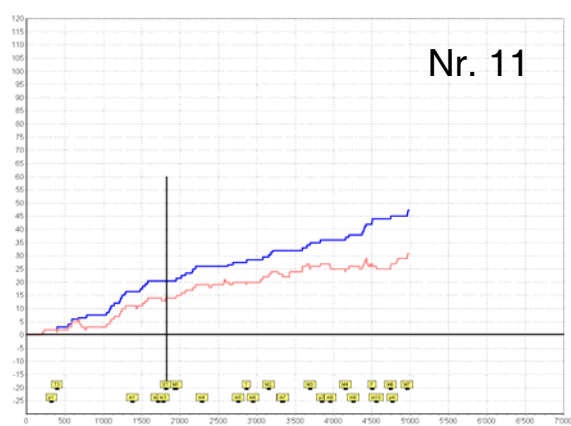
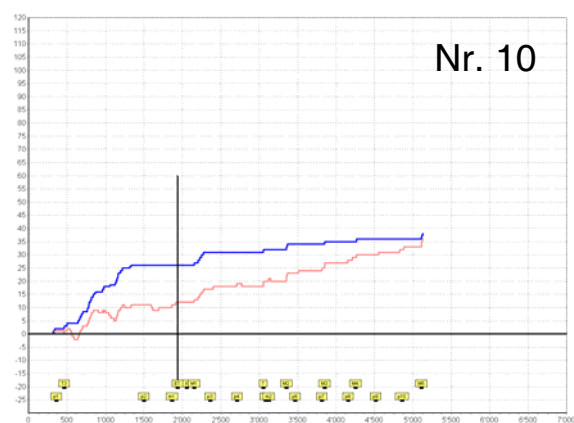
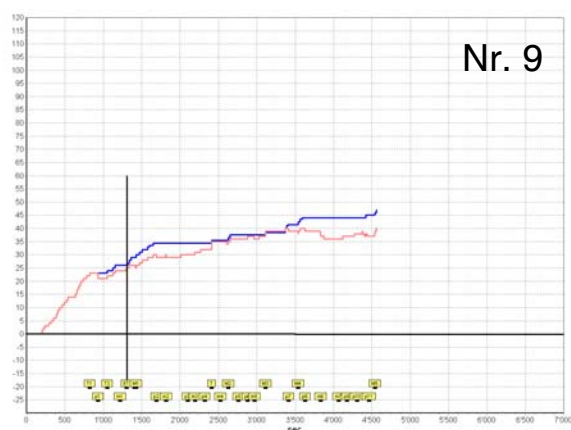
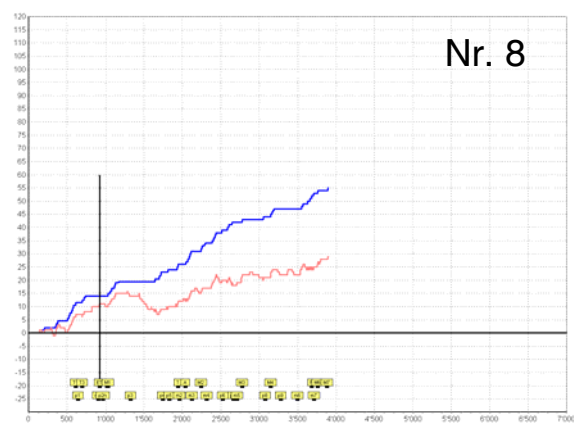
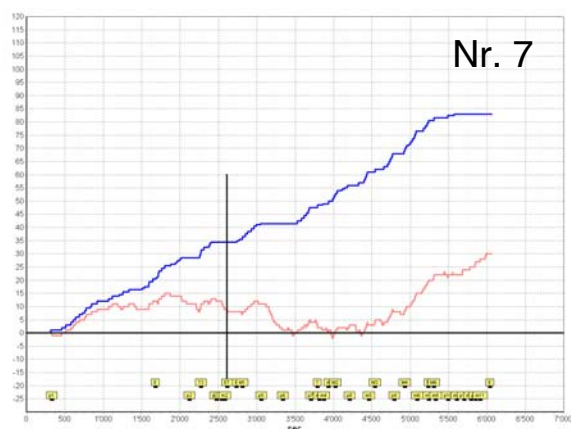


Abbildung 18 (Fortsetzung): Darstellung der Zeitreihen für die Teilstichprobe 1.

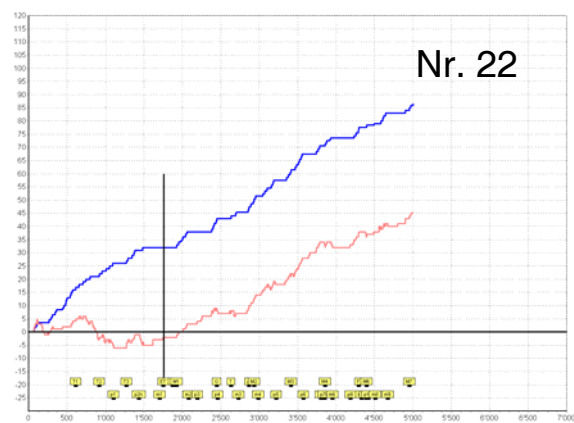
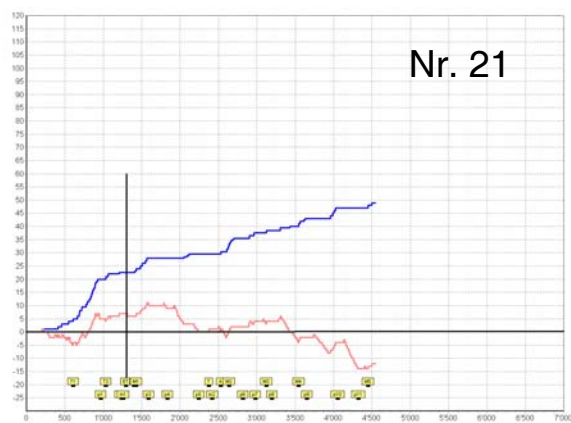
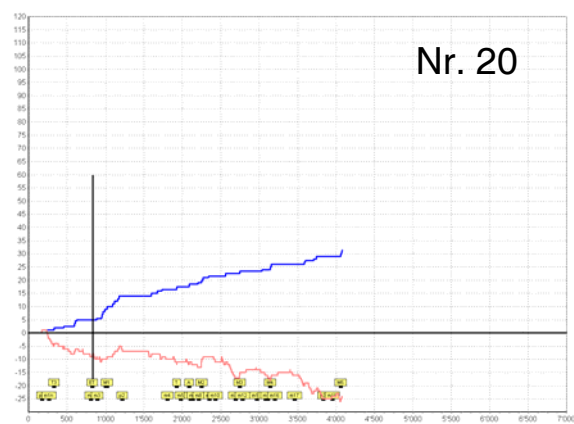
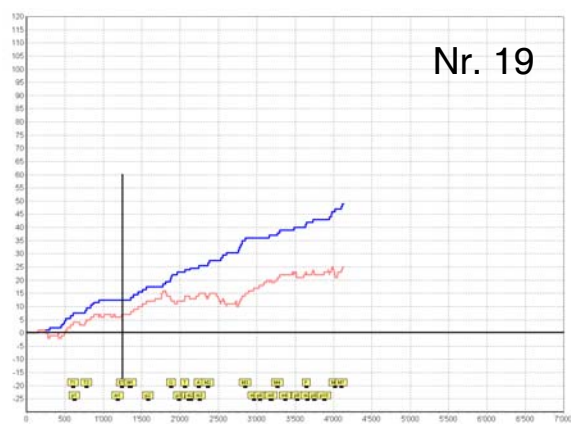
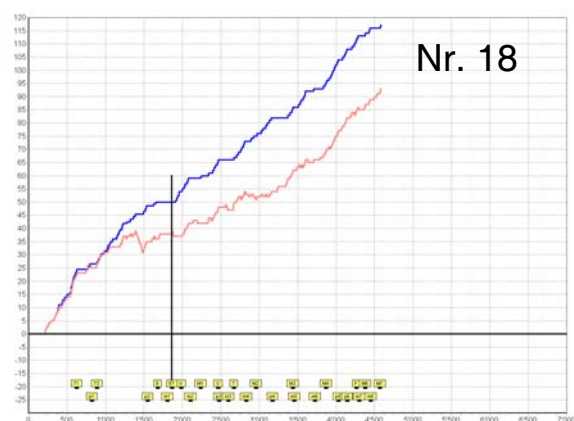
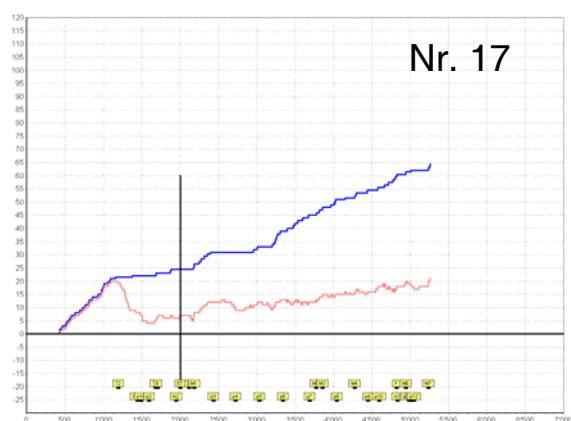
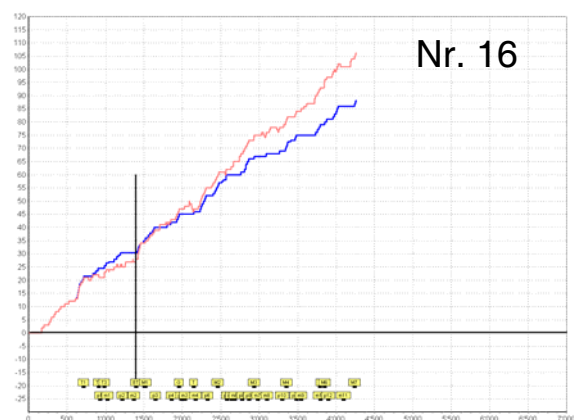
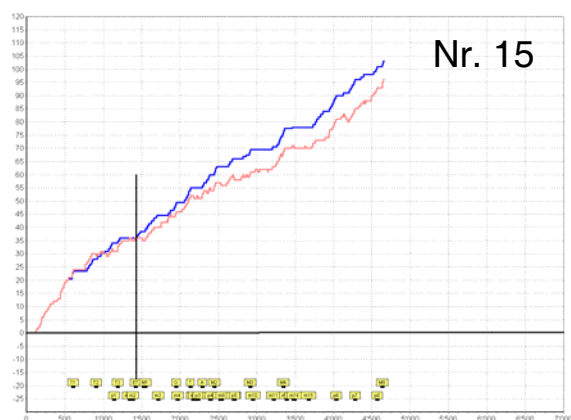


Abbildung 18 (Fortsetzung): Darstellung der Zeitreihen für die Teilstichprobe 1.

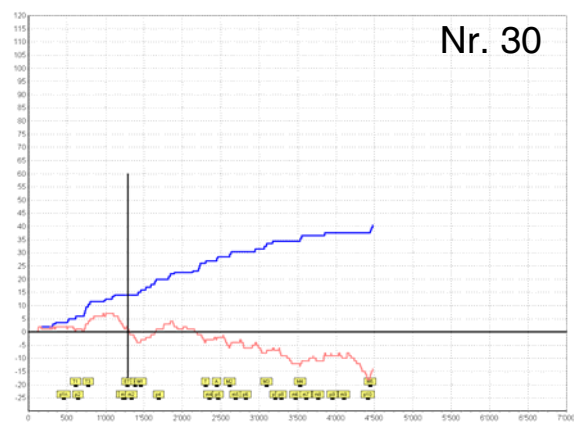
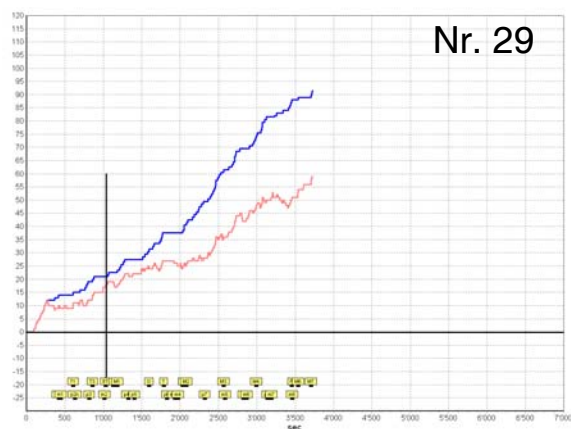
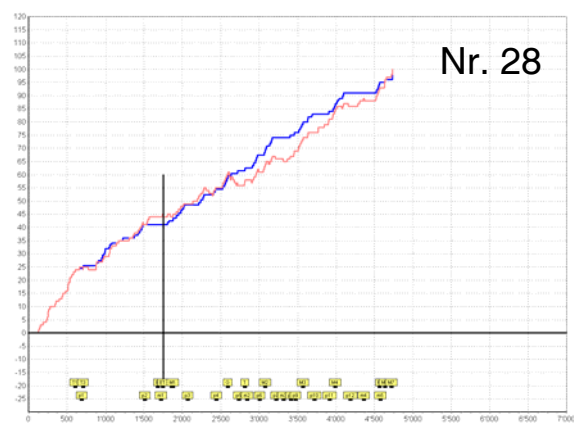
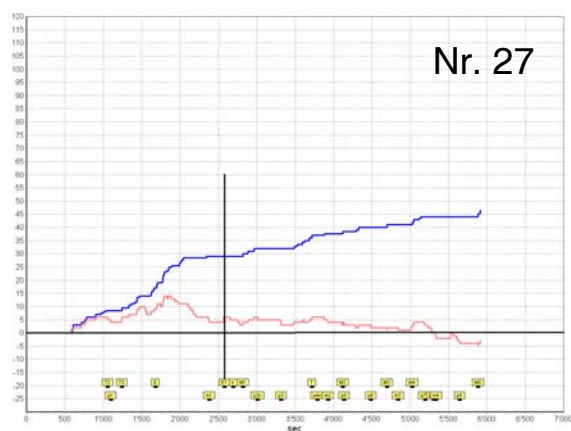
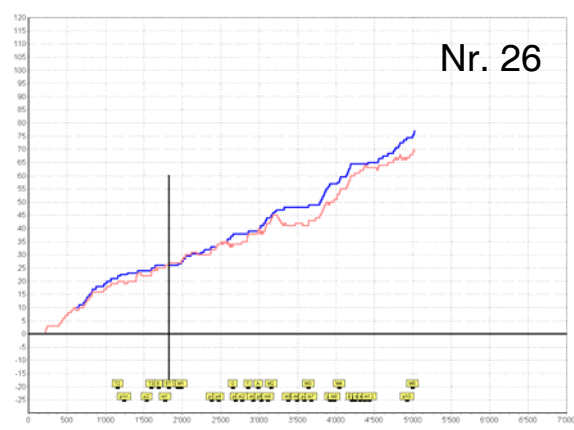
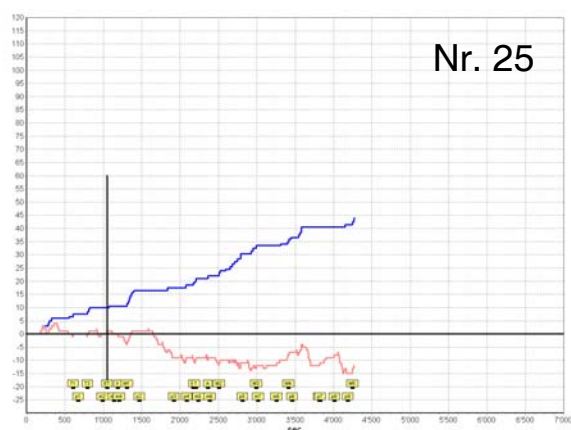
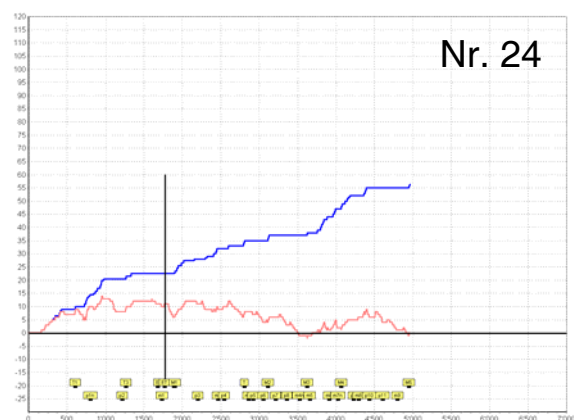
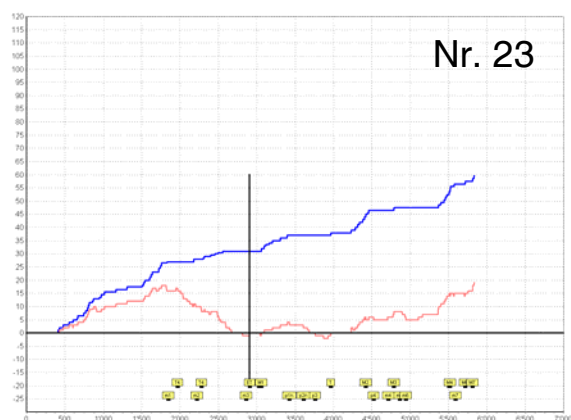


Abbildung 18 (Fortsetzung):

Darstellung der Zeitreihen für die Teilstichprobe 1.

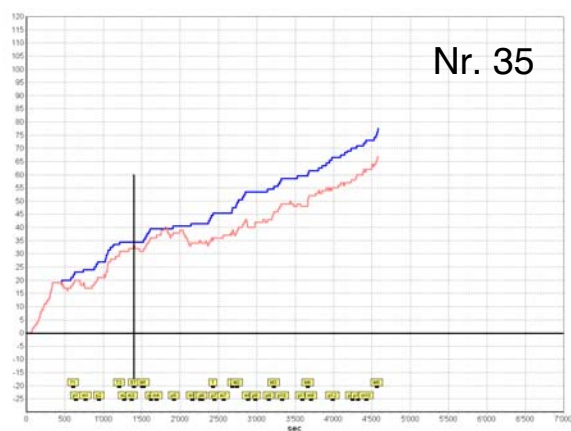
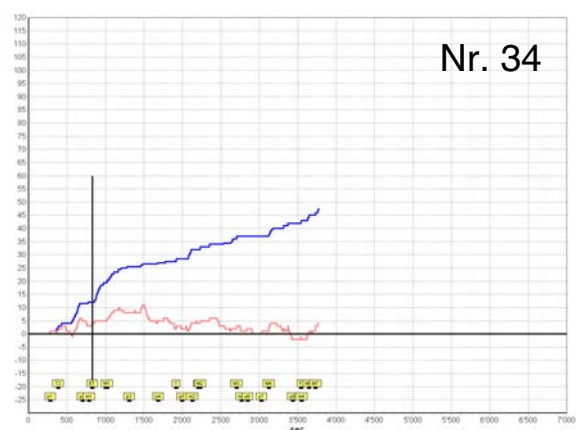
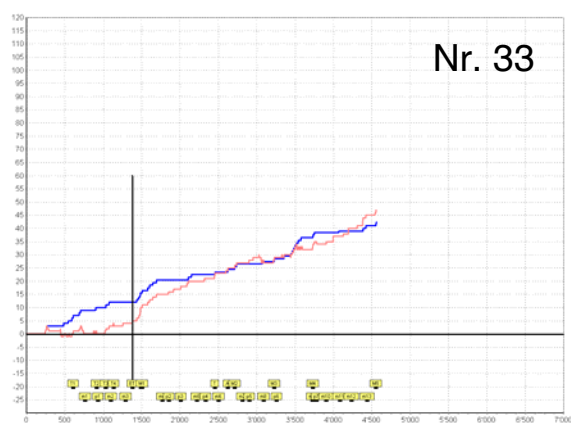
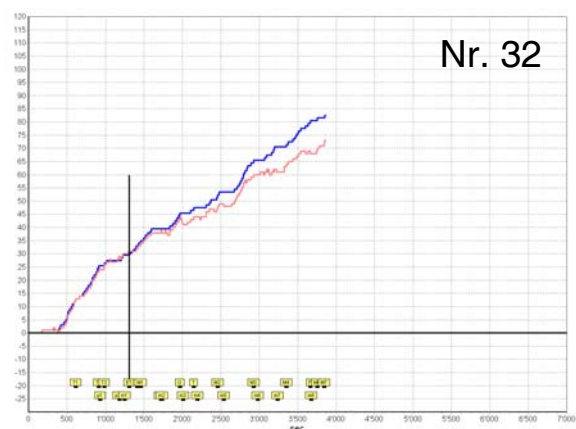
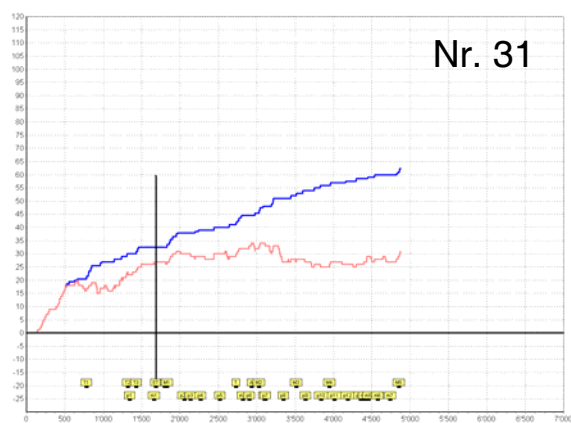


Abbildung 18 (Fortsetzung): Darstellung der Zeitreihen für die Teilstichprobe 1.

*Ergänzungen zu Kapitel 4.2.2**Tabelle 12:* Verteilungskennwerte des Informationseröffnenden Handelns zu ausgewählten Spielzeitpunkten.

		T3	X	T	M2	M3	M4	Ende
n		34	35	34	35	35	35	35
Minimum		2.0	9.0	17.5	20.0	23.5	25.0	31.5
Maximum		35.0	54.0	67.0	75.0	86.0	95.0	117.0
Mittelwert		18.2	26.2	38.0	42.1	48.9	54.3	65.6
Standardabweichung		9.8	10.3	12.2	13.6	16.0	17.9	21.9
Perzentile	25	9.5	19.5	28.5	33.0	37.0	38.0	47.5
	50	20.3	26.0	37.3	41.5	45.0	52.5	62.5
	75	25.5	32.5	45.6	50.5	61.5	70.0	82.5

Tabelle 13: Verteilungskennwerte des Adäquaten Handelns zu ausgewählten Spielzeitpunkten.

		T3	X	T	M2	M3	M4	Ende
n		34	35	34	35	35	35	35
Minimum		-18.0	-13.0	-11.0	-12.0	-15.0	-17.0	-34.0
Maximum		33.0	44.0	58.0	64.0	75.0	85.0	106.0
Mittelwert		10.2	13.3	18.1	19.8	23.1	26.3	32.8
Standardabweichung		11.9	14.6	19.7	21.7	25.8	28.7	37.6
Perzentile	25	2.0	5.0	1.8	1.0	3.0	3.0	1.0
	50	6.5	10.0	13.0	14.0	22.0	22.0	29.0
	75	22.3	27.0	35.3	38.0	43.0	48.0	62.0

Tabelle 14: Interkorrelationen[†] zwischen den verschiedenen Spielzeitpunkten für das Informationseröffnende und Adäquate Handeln.

		T3	X	T	M2	M3	M4
X	Info	.70					
	Ad	.94					
	n	34					
T	Info	.70	.96				
	Ad	.84	.95				
	n	33	34				
M2	Info	.71	.95	.99			
	Ad	.82	.93	.99			
	n	34	34	34			
M3	Info	.68	.88	.94	.97		
	Ad	.77	.90	.97	.98		
	n	34	34	34	35		
M4	Info	.71	.85	.92	.95	.99	
	Ad	.74	.87	.96	.97	.99	
	n	34	34	34	35	35	
Ende	Info	.72	.81	.89	.92	.96	.98
	Ad	.69	.81	.90	.92	.94	.96
	n	34	34	34	35	35	35

[†] Bemerkung: Produkt-Moment-Korrelationen. Alle Korrelationen sind hoch signifikant.

Ergänzungen zu Kapitel 4.4.1

Tabelle 15: Medianwerte für die Leistungskennwerte in Abhängigkeit von den Prozessprofilen für das Informationseröffnende und Adäquate Handeln (n = 35).

	Informationseröffnendes Handeln			Adäquates Handeln			
	1 (n = 7)	2 (n = 13)	3 (n = 15)	1 (n = 6)	2 (n = 7)	3 (n = 7)	4 (n = 15)
p_metall	36.8	21.1	5.3	44.7	15.8	5.3	5.3 †
p_met_s7	66.7	22.2	0.0	72.3	11.1	0.0	0.0 †
p_met_r9	30.0	10.0	0.0	25.0	30.0	0.0	10.0 †
p_ow_m †	70.8	58.5	54.6	70.2	70.8	67.2	38.7 †
p_scan_v	73.6	67.9	55.9	76.5	70.5	64.3	50.0
p_s_mg	96.9	84.3	50.0	97.0	96.6	65.1	57.1
p_v_mg	87.0	57.1	25.0	89.4	70.0	64.3	25.0
p_s_mg79	85.7	67.5	24.5	87.6	86.7	50.0	38.1
p_v_mv79	80.0	37.0	0.0	80.5	56.0	0.0	0.0 †
p_mess15	90.0	80.0	62.5	90.0	87.5	62.5	70.0 †
Aggregiertes Mass (alle Kennwerte, n = 28)	75.3	61.3	38.3	75.4	66.4	57.0	40.9
Aggregiertes Mass (10 Kennwerte)	70.9	52.2	32.6	72.7	62.3	43.9	37.7

† Einziger Kennwert mit nicht signifikanten Gruppenunterschieden (Jonckheere-Terpstra Test).

‡ Reihenfolge der Medianwerte entspricht nicht der postulierten Trendhypothese (Profil 1 > 2 > 3 > 4).

Ergänzungen zu Kapitel 4.4.2 und Kapitel 4.4.3

Tabelle 16: Übereinstimmung zwischen neuem Auswertungsmodell und Expertenurteil in Bezug auf die Problemlösetypen in der Teilstichprobe 1 (angegeben sind die absoluten Häufigkeiten).

			Prozessprofile für das Informationseröffnende Handeln			Total
			1	2	3	
Prozessprofile für das Adäquate Handeln	1	Neue Auswertung	6	0		6
		Experte	5	1		6
		Konkordanzen	4	0		4
	2	Neue Auswertung	1	5	1	7
		Experte	1	2	1	4
		Konkordanzen	1	1	0	2
	3	Neue Auswertung		3	4	7
		Experte		4	6	10
		Konkordanzen		1	0	1
	4	Neue Auswertung		5	10	15
		Experte		5	10	15
		Konkordanzen		1	6	7
Total	Neue Auswertung		7	13	15	35
	Experte		6	12	17	35
	Konkordanzen		5	3	6	14

Tabelle 17: Übereinstimmung zwischen neuem Auswertungsmodell und Expertenurteil in Bezug auf die Problemlösetypen in der Teilstichprobe 2 (angegeben sind die absoluten Häufigkeiten).

			Prozessprofile für das Informationseröffnende Handeln			Total
			1	2	3	
Prozessprofile für das Adäquate Handeln	1	Neue Auswertung	0	1		1
		Experte	2	0		2
		Konkordanzen	0	0		0
	2	Neue Auswertung	1	2	0	3
		Experte	1	6	1	8
		Konkordanzen	0	1	0	1
	3	Neue Auswertung		5	1	6
		Experte		4	1	5
		Konkordanzen		3	0	3
	4	Neue Auswertung		5	6	11
		Experte		1	5	6
		Konkordanzen		0	4	4
Total	Neue Auswertung		1	13	7	21
	Experte		3	11	7	21
	Konkordanzen		0	4	4	8

Curriculum Vitae

Personalien

Geburtsdatum 12.11.1976
Heimatort Zürich
Zivilstand verheiratet
Kinder 1 Sohn (geb. 20.04.2004)

Ausbildung

Herbst 2007 Doktorat an der Philosophischen Fakultät der Universität Zürich
Frühling 2003 Lizentiat an der Philosophischen Fakultät der Universität Zürich
1997 – 2003 Studium der Psychologie an der Universität Zürich
Vertiefungsrichtung: Abteilung Angewandte Psychologie, Lehrstuhl Psychologische Methodenlehre (Prof. Dr. R. Hirsig)
Nebenfächer: Psychopathologie des Erwachsenenalters (Prof. Dr. H.-J. Haug) und Neurophysiologie (Prof. Dr. M.-C. Hepp-Reymond)
Lizenziatsarbeit: Ressourcen schizophrener Patienten. Validitätsprüfung des Erhebungsbogens zur Erfassung gesunder Anteile.
1996 – 1997 Studium der Medizin an der Universität Zürich (1. Propädeutikum)
Januar 1996 Matura Typus B
1983 – 1995 Kantonsschule Sargans (SG), Sekundarschule Sevelen (SG), Primarschule Sevelen (SG) und Madetswil (ZH)

Berufliche Tätigkeiten

Seit 2003 Forschungstätigkeit und Lehre als Assistentin und Doktorandin am Psychologischen Institut der Universität Zürich, Lehrstuhl Psychologische Methodenlehre (Prof. Dr. R. Hirsig)
2002 4-monatiges Praktikum im Psychologischen Dienst der Psychiatrischen Universitätsklinik Zürich (Dr. H.-M. Zöllner)
2001 – 2002 Studentische Vertretung in der Institutskonferenz des Psychologischen Instituts der Universität Zürich
1998 – 2002 15 verschiedene Semesterassistenzen/Tutorate (Statistik, Testpsychologie, Psychopathologie) am Psychologischen Institut der Universität Zürich
1999 – 2000 Mitarbeit (statistische Auswertung) beim Projekt FIT-2000 am Psychologischen Institut der Universität Zürich, Lehrstuhl Angewandte Psychologie (Prof. Dr. F. Stoll)
1996 – 1999 Verschiedene Praktika und Teilzeitanstellungen (als Praktikantin im Ärztlichen Dienst, Betreuerin und Hilfsschwester)
Psychiatrische Klinik Königsfelden (AG), Wohnheim „Neufeld“ für geistig und körperlich behinderte Erwachsene in Buchs (SG), Rehabilitationsklinik Walenstadtberg (SG)